REPORT DOCUMENTATION PAGE	2	Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
15-09-2014	Ph.D. Dissertation	-

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Large Deviations and Quasipotential for finite state mean field interacting particle systems	W911NF-12-1-0222
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER 611102
6. AUTHORS Wei Wu	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES	8. PERFORMING ORGANIZATION REPORT NUMBER
Brown University Box 1929 Providence, RI 02912 -9093	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)	10. SPONSOR/MONITOR'S ACRONYM(S) ARO
U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	11. SPONSOR/MONITOR'S REPORT NUMBER(S) 61759-MA.14

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

We study a general class of mean field interacting particle systems with a finite state space. Particles evolve as exchangeable jump Markov processes, where finite collections of particles are allowed to change their state simultaneously. Such models arise naturally in statistical physics, queueing systems and communication networks literatures.
In the first part of the thesis, we establish a large deviation principle for the empirical measure process for the

15. SUBJECT TERMS

large deviations, quasipotential, mean field interacting particle systems, Markov chain approximations

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Paul Dupuis
UU	UU	UU	UU		19b. TELEPHONE NUMBER 401-863-3238

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

Report Title

Large Deviations and Quasipotential for finite state mean field interacting particle systems

ABSTRACT

We study a general class of mean field interacting particle systems with a finite state space. Particles evolve as exchangeable jump Markov processes, where finite collections of particles are allowed to change their state simultaneously. Such models arise naturally in statistical physics, queueing systems and communication networks literatures.

In the first part of the thesis, we establish a large deviation principle for the empirical measure process for the interacting particle systems. The approach is based on a variational representation for functionals of a Poisson random measure. Under an appropriate communication condition, we also prove a locally uniform large deviation principle. The main novelty is that more than one particle is allowed to change its state simultaneously, and so a standard approach to the proof based on change of measure is not possible. Along the way, we establish an LDP for jump Markov processes on the simplex, whose rates decay to zero as they approach the boundary of the domain. This result may be of independent interest.

In the second part of the thesis, we focus on the mean field interacting particle systems that only admit single particle jumps. Under the assumption that there exists a unique stationary measure, we construct a Markov chain approximation of the quasipotential function associated with the equilibrium. This is the first example of the Markov chain approximation for problems with non-quadratic running cost (but still convex in the control), which may also have singularities near the boundary.

Large Deviations and Quasipotential for finite state mean field interacting particle systems

by

Wei Wu

B.Sc., Fudan University; Shanghai, China 2009

M.Sc., Brown University; Providence, RI 2010

A dissertation submitted in partial fulfillment of the

requirements for the degree of Doctor of Philosophy

in The Division of Applied Mathematics at Brown University

PROVIDENCE, RHODE ISLAND

May 2014

This dissertation by Wei Wu is accepted in its present form
by The Division of Applied Mathematics as satisfying the
dissertation requirement for the degree of Doctor of Philosophy.

Date______________ ____________________________

Paul Dupuis, Ph.D., Advisor

Recommended to the Graduate Council

Date______________ ____________________________

Kavita Ramanan, Ph.D., Advisor

Date______________ ____________________________

Wendell Fleming, Ph.D., Reader

Approved by the Graduate Council

Date______________ ____________________________

Peter M. Weber, Dean of the Graduate School

Vitae

Wei Wu [redacted]. He obtained his B.Sc. degree in Physics from Fudan University in Shanghai in June 2009. In the fall of 2009, he started a Ph.D. program in the Division of Applied Math at Brown University, and obtained his Sc.M. degree in Applied Math in May 2010. This dissertation was defended on May 9, 2014.

In the fall of 2014, he will join the Courant Institute of Mathematical Sciences, New York University, as a Courant Instructor. He will serve as a Global Postdoctoral Fellow at the Mathematical Institute, NYU Shanghai, in the fall of 2015.

Acknowledgements

First of all, I am deeply grateful to my thesis advisor, Paul Dupuis, for his guidance, support and patience during the course of my Ph.D. studies, that changes me from a person interested in probability into a potential probabilist. He has been spending a lot time in passing his knowledge in large deviations and control theory to me, and in offering insights and suggesting approaches when most needed. With great patience he read through and revise the early drafts. He has devoted a lot effort in the topic, the structure and the details of the present thesis. I am also influenced by his rigorous attitude towards research.

Next, I would like to express my gratitude to Kavita Ramanan, who is my thesis co-advisor and the current director of graduate studies. She did a lot to help me improve my work. Over the years, she spent a lot effort in providing numerous advice, suggestions and feedbacks for my projects, that greatly helps me with writing and revising the thesis work.

I am also grateful to Richard Kenyon, who guided me into the area of probability and statistical mechanics, when I started at Brown. With his guidance and help, I completed a joint work with him.

Special thanks to Wendell Fleming, who serves as my thesis reader and spends efforts reading carefully through the manuscripts. As a much respected mathematician, he has always been kind and willing to help young people like me. I am also grateful for his career advice.

My sincere thanks to Hongjie Dong, who is both my teacher at Brown and my alumni at Fudan, and helped me a lot over the years. I also thank him for commenting on the thesis manuscripts.

I would like to take this opportunity to thank those who in various ways contribute to the completion of my Ph.D. studies. Thank Constantine Dafermos, Bjorn Sandstede, Sergei Treil and Hui Wang for service on my prelim exams. Thanks to Jean Radican for her constant support. And, I would like to thank the following friends, for their help and support in and out mathematics: Zheng Chen, Yu Gu, Qijun Hong, Adrien Kassel, Jing-Guo Lai, Xin Sun, Sirui Tan, Wei-Ying Wang, Hong Zhang and Xueyu Zhu. Your presence made my Ph.D. studies more complete.

Last, but certainly not least, I thank my parents for their love, understanding and encouragement that give me the courage to move forward.

Abstract of “ Large Deviations and Quasipotential for finite state mean field interacting particle systems ” by Wei Wu, Ph.D., Brown University, May 2014

We study a general class of mean field interacting particle systems with a finite state space. Particles evolve as exchangeable jump Markov processes, where finite collections of particles are allowed to change their state simultaneously. Such models arise naturally in statistical physics, queueing systems and communication networks literatures.

In the first part of the thesis, we establish a large deviation principle for the empirical measure process for the interacting particle systems. The approach is based on a variational representation for functionals of a Poisson random measure. Under an appropriate communication condition, we also prove a locally uniform large deviation principle. The main novelty is that more than one particle is allowed to change its state simultaneously, and so a standard approach to the proof based on change of measure is not possible. Along the way, we establish an LDP for jump Markov processes on the simplex, whose rates decay to zero as they approach the boundary of the domain. This result may be of independent interest.

In the second part of the thesis, we focus on the mean field interacting particle systems that only admit single particle jumps. Under the assumption that there exists a unique stationary measure, we construct a Markov chain approximation of the quasipotential function associated with the equilibrium. This is the first example of the Markov chain approximation for problems with non-quadratic running cost (but still convex in the control), which may also have singularities near the boundary.

Contents

List of Tables

List of Figures

Chapter One

Introduction

Markovian particle systems on finite state spaces under mean-field interactions arise in many different contexts. They may appear as approximations of statistical physics models in higher dimensional lattices (for various type of spin dynamics, see [33], and references therein) and kinetic theory [25]. Recently, this type of model also appears in modeling communication networks [1], [21], [23], [38], and game theory [22]. The dynamics of these particle systems have the following common features: a) particles are exchangeable, that is, their joint distribution is invariant under permutation of their indices; b) at each time, some group of finitely many particles can switch their state simultaneously; c) the interaction between particles is global but weak, in the sense that the jump rate (of each group of particles) is a function of the initial and final configurations only of the group of particles, and the empirical measure of all particles. The precise dynamics of the Markovian $n-$particle system we consider are described in Section 2.1.

Due to the exchangeability assumption, many essential features of the state of the particle system can be captured by its empirical measure, which evolves as a jump Markov process on (a sublattice of) the unit simplex. Under mild assumptions on the jump rates, standard results on jump Markov processes (see [34]) show that the functional law of large numbers limit of the sequence of $n-$particle empirical measures is the solution to a nonlinear ordinary differential equation (ODE) on the unit simplex. The ODE also characterizes the transition probabilities of a certain "nonlinear Markov process" [26] that describes the limiting distribution of a typical particle in the system, as the number of particles goes to infinity, and is commonly referred to as the McKean-Vlasov limit. The first question we address is the sample path large deviation properties of the sequence of empirical measure processes as the number of particles tends to infinity. In the case of interacting diffusion processes, such an LDP was first established by Dawson and Gartner in [9]. The sample path large deviation principle over finite time intervals has a number of applications, including the study of metastability properties (via Freidlin-Wentzell theory [20]), or to quantify the fact that a Gibbs measure may evolve into a non-Gibbs measure (i.e., no reasonable Hamiltonian can be defined) under stochastic dynamics (which is called the Gibbs-non Gibbs transtions in [39]).

Large deviations principles for jump Markov processes are known if the jump rates are Lipschitz continuous and uniformly bounded away from zero (cf. [36]). In this case, the large deviation rate function admits an integral representation in terms of a so-called local rate function. However, this condition is not satisfied by our model. Specifically, as the empirical measure approaches the boundary of the simplex, its jump rates along certain directions converge to zero. Nevertheless, we show that (under general conditions on the jump rates), the sequence of empirical measure processes satisfies a sample path LDP with the rate function having the standard integral representation. Under mild conditions, we also establish a "locally uniform" refinement [36], which characterize the decay rate of the probabilities of hitting a convergent sequence of points. Such result is of relevance only for discrete Markov processes (and not for diffusions) and does not follow immediately from the LDP. Indeed, we provide an example (Example 3.1.26 in Section 3.1.4) where the sample path LDP holds, but its locally uniform refinement does not. The locally uniform refinement is shown in [7] to be relevant for the study of stability properties of the nonlinear ODE that describes the law of large numbers (LLN) limit. All the main results of this paper are formulated for a more general class of jump Markov processes on the simplex whose rates diminish to zero at the boundary, and the interacting particle models are obtained as a special case.

Other works that have studied large deviations for jump Markov processes with vanishing rates include [35], [30] and [3]. However, the results in [35] impose special conditions on the jump rates near the boundary, which do not apply to our model (see Appendix A). On the other hand, the method in [30] and [3] are adaptations of the argument used by Dawson and Gartner in [9], which crucially relies on the fact that the measure on path space induced by the interacting $n-$particle process is absolutely continuous with respect to that induced by n independent (non-interacting) particles, each evolving according to a time inhomogenous Markov process. This property does not hold when multiple particles jump simultaneously. Simultaneous jumps are a common feature of models used in many applications (see Example 2.4.3 and also [38] and [18, Chapter 8]).

The large deviation upper bound follows from general results in [13] (see Section 3.4). The subtlety is to prove the large deviation lower bound. Our strategy for the proof is based on a variational representation for the $n-$particle empirical measure process and a perturbation argument near the boundary. The starting point of our variational representation is a representation formula for the functionals of Poisson random measures [6], and an SDE representation of the empirical measure process in terms of a sequence of Poisson random measures. However, the state dependent nature of the jump rates leads to a somewhat complicated variational problem. We use the special structure of the SDE to simplify the representation formula. The perturbation argument takes inspiration from [15], where an LDP was established for a discrete time one dimensional Markov chain. Our model is higher dimensional, where the perturbation argument becomes substantially more intricate, and geometry comes into play. The variational representation that we establish holds more generally for jump Markov processes with bounded jump rates, and could be useful for obtaining other asymptotics.

In the situation when the law of large numbers dynamics of the Markov process has multiple attractors in a domain, the system typically spends a long time in a small neighborhood of an attractor and rarely makes an excursion away from it. After a long time and multiple excursion attempts, a large fluctuation occurs and makes one of these attempts succeed, so the system approach one of the other attractors. Such phenomena of noise induced rare jumps between attractors is called metastability.

The work of Freidlin and Wentzell establishes the connections between large deviation principle and the metastable properties of the underlying Markov process. A key notion in their framework is the so called quasipotential, which has a representation takes the form of calculus of variations (as an exit problem). The quasipotential provides important information that can be used to study transitions of the interacting particle systems between different regions of the domain. It provides the asymptotic descriptions of the likelihood of a path and the exit distribution from a domain (at the level of exponential rates) when such a transition happens.

In most cases, one cannot obtain an explicit solution for the quasipotential, and therefore a numerical approximation is needed. A standard method for the construction of approximations in optimal control problems is the Markov chain approximation method. This method is based on directly approximating the controlled process that appear in the variational problem by a finite state controlled Markov chain, and define an analogous minimal cost function as the approximation. This method has been used in the context of [28], [29], [4], [16] and so on, and is especially useful when the corresponding PDE characterization (Hamilton-Jacobi-Bellman equations) is not well defined or difficult to work with.

There are many standard cost structure for which the convergence of the Markov chain approximation as the discretization goes to zero is well known. However, our cost structure is novel, in the sense that the linear quadratic structure (and strict positivity) of the running cost no longer hold, and the boundary of the associated exit problem is a singleton. We adapt the method of [29] and [4] in our setting, and in an upcoming work [17] we prove a convergence result of the approximation problem.

The outline of this paper is as follows. In Chapter 2 we set up the mean field interacting particle system, and describe a few examples in the literature that fit into our framework. Chapter 3 is devoted to the proof of the large deviation principle for the empirical measure process for such interacting particle systems. Some technical details are deferred to Appendix B and C. In Chapter 4, we restrict to the mean field interacting particle system with single particle jumps. We construct Markov chain approximations of the quasipotential function associated with the stationary measure of such interacting particle systems, and give a proof of convergence. Finally, in Appendix A we briefly describe why the conditions in [35] do not apply to our large deviation problem.

Chapter Two

The Interacting Particle Systems

2.1 Model Description

Here we introduce the finite state mean-field interacting particle system. Consider an n-particle system in which the state of each individual particle takes values in the finite set $\mathcal{X} = \{1, 2, ..., d\}$. Let $P(\mathcal{X})$ denote the space of probability measures on $\mathcal{X}$. We identify $P(\mathcal{X})$ with the simplex $\mathcal{S} = \{p \in \mathbb{R}^d : p_i \geq 0, \sum_{i=1}^d p_i = 1\}$ and endow $\mathcal{S}$ with the topology induced from $\mathbb{R}^d$, so that $P(\mathcal{X})$ is equipped with the Euclidean norm $\|\cdot\|$. Define $P_n(\mathcal{X}) = \left\{\frac{1}{n}\sum_{i=1}^n \delta_{x_i} : x \in \mathcal{X}^n\right\} \subset P(\mathcal{X})$, where δ_x represents the Dirac mass at x. Then $P_n(\mathcal{X})$ can be similarly identified with the lattice $\mathcal{S}_n = \mathcal{S} \cap \frac{1}{n}\mathbb{Z}^d$.

For each $i = 1, ..., n$, let $X^{i,n}(t)$ be the state of the i^{th} particle at time t. We assume that the sequence of processes $X^n(\cdot) = \{X^n(t) = (X^{1,n}(t), ..., X^{n,n}(t)), t \geq 0\}$, $n \in \mathbb{N}$, are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Each $X^n(\cdot)$ evolves as a càdlàg, $\mathcal{X}^n$-valued jump Markov process. The associated empirical measure is denoted by

$$\mu^n(t, \omega) = \frac{1}{n}\sum_{i=1}^n \delta_{X^{i,n}(t,\omega)}, \quad t \geq 0, \omega \in \Omega.$$

In subsequent discussions, we will often suppress the dependence on ω.

The possible transitions of X^n are as follows. It is assumed that there exists $K \in \mathbb{N}$ such that at most K particles jump simultaneously. When $K = 1$, then almost surely at most one particle can instantaneously change its state. If the particle changes its state from i to j, for some $i, j \in \mathcal{X}$, the transition rate is assumed to be $\Gamma_{ij}^n(\mu^n(t))$, where $\{\Gamma^n(x), x \in \mathcal{S}\}$ is a family of nonnegative $d \times d$ matrices, and to be consistent with a convention used when $K > 1$ we set $\Gamma_{ii}^n(x) = 0$. In the case of general K, for each $1 \leq k \leq K$, an ordered collection of k particles among all possible ordered $k-$tuples of the $n-$particle system can simultaneously change its configuration from $\mathbf{i} = (i_1, .., i_k) \in \mathcal{X}^k$ to $\mathbf{j} = (j_1, ..., j_k) \in \mathcal{X}^k$, where $i_l \neq j_l$, for $l = 1, ..., k$. Note that it is possible that several particles may be in the same state. Let

$$\mathcal{J}^k = \left\{ (\mathbf{i}, \mathbf{j}) \in \mathcal{X}^k \times \mathcal{X}^k : i_l \neq j_l \text{ for } l = 1, ..., k \right\}$$

be the collection of all possible pairs of initial and final configurations for an ordered $k-$tuple of particles. At time t, the rate of a simultaneous jump of a $k-$tuple from $\mathbf{i} \in \mathcal{X}^k$ to $\mathbf{j} \in \mathcal{X}^k$ is given by $\Gamma^{k,n}_{\mathbf{ij}} (\mu^n (t))$, where $\left\{ \Gamma^{k,n} (x) , x \in \mathcal{S} \right\}$ is a family of nonnegative $d^k \times d^k$ matrices with $\Gamma^{k,n}_{\mathbf{ij}} (x) \equiv 0$ if $(\mathbf{i}, \mathbf{j}) \notin \mathcal{J}^k$. We also assume the jump rate is independent of the ordering of the particles: if S_k denotes the group of permutations on $\{1, ..., k\}$, then

$$\Gamma^{k,n}_{\mathbf{ij}} (x) = \Gamma^{k,n}_{\sigma(\mathbf{i})\sigma(\mathbf{j})} (x) , \text{ for any } n \in \mathbb{N}, k = 1, ..., K, x \in \mathcal{S} \text{ and } \sigma \in S_k. \tag{2.1.1}$$

2.2 Dynamics of the Empirical Measure Process

If the initial condition is exchangeable, then it is clear that the $n-$particle system described above is exchangeable, and thus its state can be described by the empirical measure μ^n.

The empirical measure process $\{\mu^n (t) , t \geq 0\}$ is a càdlàg jump Markov process that takes values in $\mathcal{S}_n$. We now identify its generator $\mathcal{L}_n$. Let $\{e_i, i = 1, ..., d\}$ represent the standard basis of $\mathbb{R}^d$. When $K = 1$, the possible jump directions of $\mu^n (\cdot)$ lie in the set $\frac{1}{n}\mathcal{V}_1$, where $\mathcal{V}_1 = \{e_j - e_i, (i, j) \in \mathcal{J}^1\}$. Moreover, the number of particles in state i when the empirical measure is equal to $x \in \mathcal{S}_n$ is nx_i. Hence, the jump rate of $\mu^n(\cdot)$ in the direction $\frac{1}{n}(e_j - e_i)$ is $nx_i\Gamma^{1,n}_{ij}(x)$, and $\mathcal{L}_n$ takes the form

$$\mathcal{L}_n(f)(x) = n \sum_{(i,j)\in\mathcal{J}^1} x_i \Gamma^{1,n}_{ij}(x) \left[f\left(x + \frac{1}{n}(e_j - e_i)\right) - f(x) \right] \tag{2.2.1}$$

for any bounded function $f : \mathcal{S}_n \mapsto \mathbb{R}$.

In the general case when $K \in \mathbb{N}$, for $1 \le k \le K$, $\mathbf{i} = \{i_1, ..., i_k\} \in \mathcal{X}^k$, $n \in \mathbb{N}$ and $x \in \mathcal{S}_n$, let $y \in \mathcal{X}^n$ be a collection of n particles whose empirical measure is x. Define $A_k(n, \mathbf{i}, x)$ to be the number of ordered subsets of $\{1, ..., n\}$ of size k, with the property that if $(j_1, \ldots, j_k)$ is the subset, then $(y_{j_1}, \ldots, y_{j_k}) = (i_1, \ldots, i_k)$, i.e., the l^{th} particle in the subset is in state i_l. Then $A_k(n, \mathbf{i}, x)$ takes the form

$$A_k(n, \mathbf{i}, x) = n^k \prod_{l=1}^{k} x_{i_l} + O\left(n^{k-1}\right), \tag{2.2.2}$$

where the error term is non-zero when the states $\{i_l\}_{l=1}^k$ are not all distinct.

For $k = 1, ..., K$ and $\mathbf{i} \in \mathcal{X}^k$, denote $e_{\mathbf{i}} = \sum_{l=1}^k e_{i_l}$, and define $\mathcal{V} = \cup_{k=1}^K \mathcal{V}_k$, where

$$\mathcal{V}_k = \left\{ e_{\mathbf{j}} - e_{\mathbf{i}} :\ (\mathbf{i}, \mathbf{j}) \in \mathcal{J}^k \right\}.$$

We call $v = e_{\mathbf{j}} - e_{\mathbf{i}}$ the jump direction associated with $(\mathbf{i}, \mathbf{j})$.

Lemma 2.2.1. *The generator of the Markov process $\{\mu^n(\cdot)\}$ is given by*

$$\mathcal{L}_n(f)(x) = n \sum_{k=1}^{K} \sum_{(\mathbf{i},\mathbf{j})\in\mathcal{J}^k} \alpha^{k,n}_{\mathbf{ij}}(x) \left[f\left(x + \frac{1}{n}e_{\mathbf{j}} - \frac{1}{n}e_{\mathbf{i}}\right) - f(x) \right] \tag{2.2.3}$$

for any bounded function $f : \mathcal{S}_n \mapsto \mathbb{R}$, with

$$\alpha^{k,n}_{\mathbf{ij}}(x) = \frac{1}{n(k!)} A_k(n, \mathbf{i}, x)\, \Gamma^{k,n}_{\mathbf{ij}}(x), \ \ x \in \mathcal{S}_n. \tag{2.2.4}$$

Alternatively, the generator can be written as

$$\mathcal{L}_n(f)(x) = n \sum_{v \in \mathcal{V}} \lambda_v^n(x) \left[f\left(x + \frac{1}{n} v \right) - f(x) \right], \tag{2.2.5}$$

where

$$\lambda_v^n(x) = \sum_{k=1}^{K} \sum_{\substack{(\mathbf{i},\mathbf{j}) \in \mathcal{J}^k: \\ e_{\mathbf{j}} - e_{\mathbf{i}} = v}} \alpha_{\mathbf{ij}}^{k,n}(x). \tag{2.2.6}$$

Proof. Define an equivalence relation $\sim$ on $\mathcal{J}^k$ as follows: for $(\mathbf{i}_1, \mathbf{j}_1), (\mathbf{i}_2, \mathbf{j}_2) \in \mathcal{J}^k$, $(\mathbf{i}_1, \mathbf{j}_1) \sim (\mathbf{i}_2, \mathbf{j}_2)$ if and only if there exists $\sigma \in S_k$ such that $\sigma(\mathbf{i}_1) = \mathbf{i}_2$, $\sigma(\mathbf{j}_1) = \mathbf{j}_2$. Let $[\mathbf{i}, \mathbf{j}]$ denote the equivalence class containing $(\mathbf{i}, \mathbf{j})$, $\left[\mathcal{J}^k\right]$ denote the collection of equivalence classes, and define $S_k[\mathbf{i}, \mathbf{j}] = \{\sigma \in S_k : \sigma(\mathbf{i}) = \mathbf{i},\, \sigma(\mathbf{j}) = \mathbf{j}\}$. Since the particles are assumed indistinguishable, when $(\mathbf{i}_1, \mathbf{j}_1) \sim (\mathbf{i}_2, \mathbf{j}_2)$, the jump associated with $(\mathbf{i}_1, \mathbf{j}_1)$ coincides with the jump associated with $(\mathbf{i}_2, \mathbf{j}_2)$. Therefore, given $(\mathbf{i}, \mathbf{j}) \in \mathcal{J}^k$, the number of distinguishable ordered $k-$tuple transitions from configuration $\mathbf{i}$ to $\mathbf{j}$ is equal to $\frac{A_k(n,\mathbf{i},x)}{|S_k[\mathbf{i},\mathbf{j}]|}$. By the permutation symmetry (2.1.1), we can set $\Gamma_{[\mathbf{i},\mathbf{j}]}^{k,n}(\cdot) = \Gamma_{\mathbf{ij}}^{k,n}(\cdot)$, and the generator of the Markov process $\{\mu^n(\cdot)\}$ is given by

$$\mathcal{L}_n(f)(x) = \sum_{k=1}^{K} \sum_{[\mathbf{i},\mathbf{j}] \in \left[\mathcal{J}^k\right]} \frac{A_k(n, \mathbf{i}, x)}{|S_k[\mathbf{i}, \mathbf{j}]|} \Gamma_{[\mathbf{i},\mathbf{j}]}^{k,n}(x) \left[f\left(x + \frac{1}{n} e_{\mathbf{j}} - \frac{1}{n} e_{\mathbf{i}} \right) - f(x) \right],\ x \in \mathcal{S}_n \tag{2.2.7}$$

for any bounded function $f : \mathcal{S}_n \mapsto \mathbb{R}$.

An alternative way is to write the generator (2.2.7) as a sum over $\mathcal{J}^k$ rather than over $\left[\mathcal{J}^k\right]$. Noting that $|[\mathbf{i}, \mathbf{j}]| = \frac{|S_k|}{|S_k[\mathbf{i},\mathbf{j}]|} = \frac{k!}{|S_k[\mathbf{i},\mathbf{j}]|}$, we can rewrite (2.2.7) as in (2.2.3). $\square$

We will state and prove our results in the formulation of Markov processes with generator (2.2.5); the precise implications for the original $n-$particle system are described in Section 3.1.4.

2.3 The Law of Large Numbers Limit

The functional law of large numbers (LLN) result for general interacting jump processes was established in [34] (see also [27]). Since the properties of the law of large numbers trajectory will be used in the large deviation proof, for completeness we present a proof for our interacting particle system (Theorem 2.3.2) in Section 3.3.3.

The following condition is used to obtain a unique law of large numbers limit.

Condition 2.3.1. *For each $v \in \mathcal{V}$, there is Lipschitz continuous $\lambda_v : \mathcal{S} \to \mathbb{R}$ such that $\lambda_v^n(\cdot)$ converges uniformly to $\lambda_v(\cdot)$ on $\mathcal{S}$.*

We also define

$$M = \sup_{v \in \mathcal{V}, x \in \mathcal{S}} \lambda_v(x) < \infty. \tag{2.3.1}$$

We now state a well known law of large numbers result for the sequence of processes $\{\mu^n\}_{n \in \mathbb{N}}$ [34].

Theorem 2.3.2. *Assume the family of jump rates $\{\lambda_v^n(\cdot) : v \in \mathcal{V}, n \in \mathbb{N}\}$ satisfy Condition 2.3.1. Also, assume $\mu^n(0)$ converges in probability to $\mu_0 \in P(\mathcal{X})$ as n tends to infinity. Then $\{\mu^n(\cdot)\}_{n \in \mathbb{N}}$ converges (uniformly on compact time intervals) in probability to $\mu(\cdot)$, where $\mu(\cdot)$ is the unique solution to the nonlinear Kolmogorov forward equation*

$$\dot{\mu}(t) = \sum_{v \in \mathcal{V}} v \lambda_v(\mu(t)), \quad \mu(0) = \mu_0. \tag{2.3.2}$$

We now describe the LLN limit for the interacting particle systems. We first state an assumption on the rates $\{\Gamma_{\mathbf{ij}}^{k,n}(\cdot), (\mathbf{i},\mathbf{j}) \in \mathcal{J}^k, k = 1, ..., K, n \in \mathbb{N}\}$ that implies Condition 2.3.1.

Condition 2.3.3. *For every $k = 1, ..., K$ and $(\mathbf{i},\mathbf{j}) \in \mathcal{J}^k$, there is $\Gamma_{\mathbf{ij}}^{k} : \mathcal{S} \to \mathbb{R}$ such that for every $x \in \mathcal{S}$ and sequence $x_n \in \mathcal{S}_n$, $n \in \mathbb{N}$, such that $\lim_{n\to\infty} x_n = x$, $\Gamma_{\mathbf{ij}}^{k}(x) = \lim_{n\to\infty} n^{k-1}\Gamma_{\mathbf{ij}}^{k,n}(x_n)$, and the family of rate matrices $\{\Gamma_{\mathbf{ij}}^{k,n}(\cdot)\}$ satisfies the permutation symmetry (2.1.1). Also, each $\alpha_{\mathbf{ij}}^{k}(\cdot)$ is Lipschitz continuous, where*

$$\alpha_{\mathbf{ij}}^{k}(x) = \lim_{n\to\infty} \alpha_{\mathbf{ij}}^{k,n}(x) = \frac{1}{k!}\prod_{l=1}^{k} x_{i_l}\Gamma_{\mathbf{ij}}^{k}(x). \tag{2.3.3}$$

Given a Markov process with generator (2.2.3), one can always define a single particle jump process, with the jump rate matrix

$$\Gamma_{ij}^{n,\mathrm{eff}}(x) = \sum_{k=1}^{K}\sum_{(\mathbf{i},\mathbf{j})\in\mathcal{J}^k}\sum_{l=1}^{k}\frac{\alpha_{\mathbf{ij}}^{k,n}(x)}{x_i}1_{\{i=i_l,j=j_l\}}, \qquad x \in \mathcal{S} \tag{2.3.4}$$

for $(i,j) \in \mathcal{J}^1$ and $n \in \mathbb{N}$. When $x_i = 0$, $\alpha_{\mathbf{ij}}^{k,n}(x)/x_i$ is understood as the pointwise limit of $\alpha_{\mathbf{ij}}^{k,n}(y)/y_i$, when y lies in the relative interior of $\mathcal{S}$ and $y \to x$ in the Euclidean norm. Here the superscript "eff" stands for "effective," and indicates that the single jump process will have the same LLN limit (2.3.2). Indeed, from Condition 2.3.3, it is clear that for each $x \in \mathcal{S}$, $\Gamma_{ij}^{n,\mathrm{eff}}(x)$ converges, as $n \to \infty$, to

$$\Gamma_{ij}^{\mathrm{eff}}(x) = \sum_{k=1}^{K}\sum_{(\mathbf{i},\mathbf{j})\in\mathcal{J}^k}\sum_{l=1}^{k}\left(\prod_{\substack{r=1\\ r\neq l}}^{k} x_{i_r}\right)\Gamma_{\mathbf{ij}}^{k}(x)\,1_{\{i=i_l,j=j_l\}}. \tag{2.3.5}$$

Moreover, if Condition 2.3.3 holds, then the jump rates $\{\lambda_v^n\}_{v\in\mathcal{V}}$ defined by (2.2.6) satisfy Condition 2.3.1, with

$$\lambda_v(x) = \sum_{k=1}^{K} \sum_{\substack{(\mathbf{i},\mathbf{j})\in\mathcal{J}^k: \\ e_{\mathbf{j}}-e_{\mathbf{i}}=v}} \alpha_{\mathbf{ij}}^k(x), \tag{2.3.6}$$

However, as elaborated in Example 3.1.26, the large deviation behavior of these two systems can be quite different.

2.4 Examples

The particle systems that we describe naturally occur in a wide range of areas, including statistical mechanics (Curie-Weiss model), graphical models and algorithms, networks and queueing systems (rerouting, loss networks). We present a few illustrative examples below.

Example 2.4.1. *Opinion dynamics of the Curie-Weiss model [11]. This is a mean field model on a complete graph. Let $d = 2$, and as before, let n be the number of individuals and $X^{i,n}(t) \in \{-1, 1\}$ denote the opinion of the i^{th} individual at time t. At time 0, each individual takes opinions $X^{i,n}(0)$ independently and uniformly at random. Each individual has an i.i.d Poisson clock of rate 1. If the clock of individual i rings at time t, he/she computes the opinion imbalance*

$$M^{(i)} = \sum_{j\neq i} X^{j,n}(t-),$$

and changes opinion with probability

$$\mathbb{P}_{flip}\left(X^{i,n}(t)\right) = \begin{cases} \exp\left(-2\beta\left|M^{(i)}\right|/n\right) & \text{if } M^{(i)}X^{i,n}(t-) > 0 \\ 1 & \text{otherwise.} \end{cases},$$

The empirical measure process μ^n only takes jumps of the form $\mathcal{V} = \{e_j - e_i, i, j \in \{-1, 1\}\}$. It satisfies Condition 2.3.3 with $\Gamma = \Gamma^1$, and

$$\Gamma_{1,-1}(\mu) = \begin{cases} \exp(-2\beta(\mu_1 - \mu_{-1})) & \text{if } \mu_1 - \mu_{-1} > 0 \\ 1 & \text{otherwise.} \end{cases},$$

$$\Gamma_{-1,1}(\mu) = \begin{cases} \exp(-2\beta(\mu_{-1} - \mu_1)) & \text{if } \mu_1 - \mu_{-1} < 0 \\ 1 & \text{otherwise.} \end{cases},$$

and $\Gamma_{1,1}(\mu) = \Gamma_{-1,-1}(\mu) = 0$.

A generalization of the above example is the Curie-Weiss-Potts model with Glauber dynamics, which we described below, the mixing time of which has interesting phase tranistion properties (see [31]).

Example 2.4.2. *Glauber dynamics of Curie-Weiss-Potts model. Fix $n > 0$, $X^{i,n}$ are interpreted as spins that take value in $\mathcal{X} = \{1, ..., d\}$. Let $X = \{X^{1,n}, ..., X^{n,n}\}$. Given $\beta > 0$, the Curie-Weiss distribution is a probability measure on $\mathcal{X}^n$, given by*

$$\nu_n(X) = Z_{\beta,n}^{-1} \exp\left(-\frac{\beta}{n}\sum_{i=1}^{n}\sum_{j=i+1}^{n} 1_{\{X^{i,n}=X^{j,n}\}}\right),$$

where $Z_{\beta,n}$ is the normalizing constant. The continuous time Glauber dynamics $(X_t)_{t\geq 0}$ for the spin flips is defined as follows. Each particle has an i.i.d Poisson clock of rate 1 attached to it. Assume the clock at particle i rings at time t, we set the new value of $X^{i,n}$ to be

$$k \in \mathcal{X} \quad \text{with probability } \nu_n\left(X^{i,n} = k | X^{j,n} = X^{j,n}(t-), \forall j \neq i\right)$$

The empirical measure μ^n is interpreted as the magnitization, which evolves as a

Markov process with generator (2.2.1), with

$$\Gamma_{ij}^{1,n}(x) = \frac{\exp\left(-\frac{\beta}{n}(nx_i - 1)\right)}{\sum_{k=1}^{d}\exp\left(-\frac{\beta}{n}(nx_k - 1)\right)}, \ \text{for } x \in \mathcal{S}_n.$$

Therefore, for $x \in \mathcal{S}$,

$$\alpha_{ij}^{1}(x) = \lim_{n\to\infty} x_i\Gamma_{ij}^{1,n}(x) = x_i\frac{\exp(-\beta x_i)}{\sum_{k=1}^{d}\exp(-\beta x_k)}.$$

Simultaneous jump models arise naturally in the context of communication networks. We now provide one such example, from [21]. Many more can be found in [38], [33] and [24].

Example 2.4.3. *Alternative rerouting networks [21]. Consider a network that consists of n links, each with finite capacity C. Let $\mathcal{X} = \{0, ..., C\}$ and let $X^{i,n}$ denote the number of customers using link i. At each time t, customers or packets arrive to each link as a Poisson process with rate $\lambda > 0$. If a packet arrives to a link with spare capacity, then it is accepted to the link and occupies one unit of capacity for an exponentially distributed time with mean one. On the other hand, if a packet arrives to a link that is fully occupied, two other links are choosen uniformly at random from amongst the remaining $n-1$ links. If both chosen links have a unit of spare capacity available, the packet occupies one unit of capacity on each of the two links, for two independent, exponential clocks with mean one. Otherwise, the packet is lost. This model seeks to understand the impacts by allowing alternative routes that occupy a greater number of resources.*

The empirical measure process μ^n is a jump Markov process with jump rates *summarized as follows: for any* $i, j \in \mathcal{X}$*:*

$$\mu^n \to \begin{cases} \mu^n + \frac{1}{n}\left(e_{i+1} - e_i\right) & \text{at rate } n\lambda\mu_i^n & 0 \le i \le C-1, \\ \mu^n + \frac{1}{n}\left(e_{i-1} - e_i\right) & \text{at rate } ni\mu_i^n & 1 \le i \le C, \\ \mu^n + \frac{1}{n}\left(e_{i+1} - e_i + e_{j+1} - e_j\right) & \text{at rate } 2\lambda\frac{n^3\mu_c^n\mu_i^n\mu_j^n}{(n-1)(n-2)} & 0 \le i \ne j \le C-1, \\ \mu^n + \frac{2}{n}\left(e_{i+1} - e_i\right) & \text{at rate } \lambda\frac{n^2\mu_c^n\mu_i^n\left(n\mu_i^n - 1\right)}{(n-1)(n-2)} & 0 \le i \le C-1. \end{cases}$$

This model satisfies Condition 2.3.3 with $K = 2$*, and with matrix entries*

$$\Gamma^1_{i,i+1}(\mu) = \lambda, \quad \Gamma^1_{i,i-1}(\mu) = i, \quad \Gamma^2_{(i,j),(i+1,j+1)}(\mu) = \lambda\mu_c$$

and zero otherwise. By (2.3.5), we can calculate the effective jump rate as

$$\Gamma^{\text{eff}}_{ij}(\mu) = \Gamma^1_{ij}(\mu) + \sum_{\substack{i'\ne i, j'\ne j \\ i'\ne j'}} 2\mu_{i'}\Gamma^2_{(i,i'),(j,j')}(\mu) + \mu_i\Gamma^2_{(i,i),(j,j)}(\mu),$$

which gives $\Gamma^{\text{eff}}_{i,i+1}(\mu) = \lambda + 2\sum_{i'\ne i}\mu_{i'}\lambda\mu_c + \mu_i\lambda\mu_c = \lambda\left[1 + \mu_c\left(2 - \mu_i\right)\right]$, $\Gamma^{\text{eff}}_{i,i-1}(\mu) = \lambda$*, and* $\Gamma^{\text{eff}}_{ij}(\mu) = 0$ *otherwise.*

Chapter Three

Large Deviations for Finite State Mean Field Interacting Particle Systems

This chapter is organized as follows. In Section 3.1 we state the main result, which is a sample path LDP for a general class of state dependent jump Markov process (with rates that are allowed to vanish) as well as its locally uniform refinement. We also describe how the general result specializes to the mean field interacting particle systems. In Section 3.2 we verify the conditions of the main results for the empirical measure processes associated with a large class of interacting particle systems. Section 3.3 establishes the variational representation for the empirical measure process, and provides an alternative proof for the functional LLN limit. Some details of the proof of the variational representations are deferred to Appendix B. The sample path large deviation upper and lower bounds are derived in Sections 3.4 and 3.6, respectively, while in Section 3.5 we study properties of the local rate function. Section 3.7 is devoted to the proof of the locally unifom LDP.

3.1 Main Results

3.1.1 The large deviation principle

In practice one is often interested in estimating the tail probability $\mathbb{P}(\mu^n(\cdot) \in A)$ for certain sets A that do not contain the law of large numbers limit $\mu(\cdot)$. This can be studied in the framework of a large deviation principle. We first establish a sample path large deviation principle for the sequence $\{\mu^n\}_{n\in\mathbb{N}}$. Asymptotics of the tail probabilities at a given time t will follow from the contraction principle. For simplicity we assume from now on that $t \in [0,1]$, while all results in this paper can be established for t in any compact time interval by the same argument.

As is explained in the introduction, a large deviation upper bound for jump-diffusion Markov processes was obtained in [13], their local rate function can be identified with ours (see Section 3.4). The more delicate part is the proof of the large deviation lower bound. Since the transition rates of μ^n may tend to zero as μ^n approaches the boundary of $\mathcal{S}$, the local rate function can approach ∞, and the probability decays superexponentially. Our analysis requires that the jump rates go to zero at most polynomially when approaching the boundary, which is sufficient for most of the applications, and is formulated in the condition below.

Condition 3.1.1. *There is a continuous function* $f_0 : [0,1] \to [0,\infty)$ *with* $f_0(0) = 0$, *and constants* $C_1, C_2 < \infty$ *such that for any* $x, y \in \mathcal{S}$ *and* $v \in \mathcal{V}$

$$\begin{aligned} C_1 \sum_{i=1}^{d} (\log x_i - \log y_i) + f_0(\|x-y\|) &\geq \log \lambda_v(x) - \log \lambda_v(y) \\ &\geq C_2 \sum_{i=1}^{d} (\log x_i - \log y_i) - f_0(\|x-y\|). \end{aligned}$$

Let $\mathrm{int}(\mathcal{S})$ denote the relative interior of $\mathcal{S}$.

Remark 3.1.2. *Condition 3.1.1 implies that on any compact subset of* $\mathrm{int}(\mathcal{S})$, $\lambda_v(\cdot)$ *is either identically zero or uniformly bounded below away from zero.*

Another ingredient that is required for the proof of the large deviation lower bound is an upper bound on the time (or, a lower bound on the speed) taken by the law of large numbers path to hit a compact subset of $\mathrm{int}(\mathcal{S})$.

Condition 3.1.3. *There exist constants* $b > 0$ *and* $D < \infty$, *such that for any* $x \in \mathcal{S}$, *the law of large numbers path* μ *starting at* x, *as defined by (2.3.2), satisfies* $\mu_i(t) \geq bt^D$ *for* $t \in [0,1]$.

For $t \in [0,1]$ and $G \subset \mathbb{R}^d$, let $D([0,t] : G)$ denote the space of càdlàg functions on $[0,t]$ which takes value in G, and let $AC([0,t] : G)$ denote the space of absolutely continuous functions on $[0,t]$ that take values in G. To avoid singularities in the large deviation analysis, we need to assume a certain communication condition. This is formulated precisely in Condition 3.1.4 below. In what follows, given $t>0$ and a path $\phi \in AC\left([0,t]:\mathcal{S}\right)$, let

$$\mathrm{Len}\left(\phi\right)=\int_0^t\left|\dot{\phi}\left(s\right)\right|ds \tag{3.1.1}$$

denote the length of the image of ϕ.

Condition 3.1.4. *There exist constants $c>0$ and $c',p<\infty$, such that for every $x\in\mathcal{S}$ and $y\in$int($\mathcal{S}$), there exist $t\in(0,1)$, $F_y\in\mathbb{N}$, and a piecewise linear path ϕ with $\phi\left(0\right)=x$, $\phi\left(t\right)=y$, such that*

i). there exist $F_{x,y}\leq F_y$, $\{v_m\}_{m=1}^{F_{x,y}}\subset\mathcal{V}$, $0=t_0<t_1<\cdots<t_{F_{x,y}}=t$, $\{U_m\}_{m=1}^{F_{x,y}}\subset\mathbb{R}^+$, such that

$$\dot{\phi}\left(s\right)=\sum_{m=1}^{F_{x,y}}U_m v_m 1_{[t_{m-1},t_m)}\left(s\right), \quad a.e.\ s\in[0,t]\,, \tag{3.1.2}$$

ii). for $m=1,...,F_{x,y}$,

$$\lambda_{v_m}\left(\phi\left(s\right)\right)\geq c\left(\min_{i=1,\ldots,d}y_i\right)^p,\ \textit{for } s\in[t_{m-1},t_m),$$

iii). $\mathrm{Len}(\phi)$ scales with the Euclidean distance between x and y: $\mathrm{Len}(\phi)\leq c'\left\|x-y\right\|.$

Definition 3.1.5. *Given $x,y\in\mathcal{S}$, a path ϕ that satisfies properties i)–iii) will be said to be a* ***communicating path*** *(with respect to the constants c,p,c', jump directions $\mathcal{V}$ and jump rates $\{\lambda_v\}_{v\in\mathcal{V}}$).*

Remark 3.1.6. *There is some flexibility in the choice of* t, $\{U_m\}_{m=1}^{F_{x,y}}$ *and* $\{t_m\}_{m=1}^{F_{x,y}}$. *By a reparametrization of the path, we can always assume* $t = 1$ *and* $U_m = U$ *for every* m. *Moreover, if* t *is not fixed, we can always choose* $U_m = 1$. *We will make these additional assumptions when convenient.*

We now introduce an easily verifiable sufficient condition on the jump rates $\{\lambda_v\}_{v\in\mathcal{V}}$ which, together with Conditions 3.1.1 and 3.1.4, implies Condition 3.1.3.

Condition 3.1.7. *There exists* $C < \infty$, *such that for any* $x \in \mathcal{S}$ *and any* $v \in \mathcal{V}$ *and* $i \in \mathcal{X}$ *that satisfy* $\langle v, e_i\rangle < 0$, $\lambda_v(x) \leq Cx_i$.

Remark 3.1.8. *For the interacting particle systems that we study,* $\lambda_v(\cdot)$ *is specified explicitly by (2.3.6) and (2.3.3), and therefore satisfies Condition 3.1.7. Also, if* $\{\lambda_v(\cdot)\}_{v\in\mathcal{V}}$ *are Lipschitz continuous, then Condition 3.1.7 can be replaced by: for any* $v \in \mathcal{V}$ *and* $i \in \mathcal{X}$ *that satisfy* $\langle v, e_i\rangle < 0$, $\lambda_v(x) = 0$ *if* $x_i = 0$.

Lemma 3.1.9. *Assume that* $\{\lambda_v(\cdot)\}_{v\in\mathcal{V}}$ *satisfies Conditions 3.1.1, 3.1.4 and 3.1.7. Then Condition 3.1.3 is satisfied with* $D = \sum_{i=0}^{F} d^i$, *for some* $F < \infty$.

The proof of Lemma 3.1.9 is deferred to Section 3.2.3.

We now state our first large deviation result. Let $\Delta^{d-1} = \{x \in \mathbb{R}^d : \sum_{i=1}^{d} x_i = 0\}$. For $x \in \mathcal{S}$ and $\beta \in \Delta^{d-1}$, define

$$L(x, \beta) = \inf_{q:\sum_{v\in\mathcal{V}} vq_v = \beta} \sum_{v\in\mathcal{V}} \lambda_v(x)\,\ell\left(\frac{q_v}{\lambda_v(x)}\right), \tag{3.1.3}$$

where q is a vector with nonnegative components, and

$$\ell(x) = \begin{cases} x\log x - x + 1 & x \geq 0, \\ \infty & x < 0, \end{cases} \tag{3.1.4}$$

is the local rate function for a standard Poisson process. For $t \in [0,1]$ and an absolutely continuous function $\gamma : [0,t] \mapsto \mathcal{S}$, define

$$I_t(\gamma) = \int_0^t L(\gamma(s), \dot{\gamma}(s))\, ds, \tag{3.1.5}$$

and in all the other cases, set $I_t(\gamma) = \infty$. We will write $I(\gamma)$ to denote $I_1(\gamma)$.

In what follows we equip $D([0,1] : \mathcal{S})$ with Skorohod topology, and let $\mathcal{B}(D([0,1] : \mathcal{S}))$ be the associated Borel sets.

Theorem 3.1.10. *Suppose the family of jump rates $\{\lambda_v(x), x \in \mathcal{S}, v \in \mathcal{V}\}$ satisfies Conditions 2.3.1, 3.1.1, 3.1.3 and 3.1.4. Also, assume that the initial conditions $\{\mu^n(0)\}_{n\in\mathbb{N}}$ are deterministic, and $\mu^n(0) \to \mu_0 \in P(\mathcal{X})$ as n tends to infinity. Then the corresponding sequence of empirical measure processes $\{\mu^n\}_{n\in\mathbb{N}}$ satisfies the sample path large deviation principle (LDP) with rate function I. Specifically, for any measurable set $A \in \mathcal{B}(D([0,1] : \mathcal{S}))$, we have the large deviation lower bound*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\mu^n \in A) \geq -\inf\{I(\gamma) : \gamma \in A^\circ, \gamma(0) = \mu_0\} \tag{3.1.6}$$

and the upper bound

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\mu^n \in A) \leq -\inf\left\{I(\gamma) : \gamma \in \overline{A}, \gamma(0) = \mu_0\right\}. \tag{3.1.7}$$

Moreover, for any compact set $K \subset \mathcal{S}$ and $M < \infty$, the set

$$\{\gamma \in AC([0,1] : \mathcal{S}) : I(\gamma) \leq M, \gamma(0) \in K\} \tag{3.1.8}$$

is compact.

The space $D([0,1] : \mathcal{S})$ is naturally equipped with Skorohod topology. However, as is shown in Lemma 2.4 of [13], to prove the large deviation result (Theorem 3.1.10), one can approximate a path in $D([0,1]:\mathcal{S})$ by a piecewise linear path, thus it suffices to consider the uniform topology. The proof of Theorem 3.1.10 is deferred to Section 3.4 and 3.6. Applying the contraction principle (see, e.g., [40]) one obtains the variational representation of the rate function of $\{\mu^n(t)\}_{n\in\mathbb{N}}$ for any $t\in[0,1]$.

Corollary 3.1.11. *Suppose the conditions of Theorem 3.1.10 hold. Then for each $t\in[0,1]$, the sequence of random variables $\{\mu^n(t)\}_{n\in\mathbb{N}}$ satisfies the LDP with rate function*

$$J_t(\mu_0,x)=\inf\{I_t(\gamma):\gamma\in D([0,1]:\mathcal{S}),\gamma(0)=\mu_0,\gamma(t)=x\} \tag{3.1.9}$$

3.1.2 A locally uniform refinement

In applications, it is often useful to estimate the probability that μ^n hits a specific point $c^n\in\mathcal{S}_n$ at some given time, where $c^n\to c\in\mathcal{S}$ as $n\to\infty$. The ordinary LDP does not imply an asymptotic rate for this hitting probability since it applies only to fixed sets, and the "moving" set $\{c^n\}$ in the present case has empty interior. To obtain such a "locally uniform" result we need a strengthening of the communication condition (Condition 3.1.4), and require local controllability of the Markov process up to the boundary of $\mathcal{S}$. It turns out that a polynomial lower bound for jump rates up to the boundary ((3.1.10) below) is sufficient for local controllability.

Definition 3.1.12. *Given $c',p,p_1<\infty$ and $c,c_1>0$, for any $x,y\in\mathcal{S}$, a path ϕ from x to y is said to be **strongly communicating**, if it is communicating (with respect to c,p,c'), and, in addition,*

iv). there exists $F<\infty$, such that $\sup_{y\in\mathcal{S}}F_y\le F$, *v). if ϕ has the representation (3.1.2), then for $m = 1, ..., F_{x,y}$,*

$$\lambda_{v_m}\left(\phi\left(s\right)\right) \geq c_1 \left(\prod_{j \in \mathcal{N}_m} \phi_j\left(s\right)\right)^{p_1}, \quad s \in [t_{m-1}, t_m), \tag{3.1.10}$$

where

$$\mathcal{N}_m = \left\{j : \langle e_j, v_m \rangle < 0\right\}. \tag{3.1.11}$$

Note that when y is on $\partial\mathcal{S}$, $\min_{i=1,\ldots,d} y_i = 0$, and therefore Condition 3.1.4.ii) is trivially satisfied. Thus a polynomial lower bound (3.1.10) is indeed a strenghthening when dealing with paths near the boundary.

Definition 3.1.13. *Given $c_1 > 0$ and $c^{'}, p_1 < \infty$, for any $x, y \in \mathcal{S}_n$, a set of points $\{\phi_0, \phi_1, ..., \phi_k\} \subset \mathcal{S}_n$ with $\phi_0 = x$, $\phi_k = y$ is called a* ***discrete strongly communicating path****, if they satisfy the following properties analogous to Definition 3.1.12:*

i). There exist $F < \infty$ (independent of x and y), $\{v_m\}_{m=1}^{F} \subset \mathcal{V}$, and $0 = t_0 < t_1 < \cdots < t_F = k$, such that

$$\phi_{s+1} - \phi_s = \frac{1}{n} v_m,\ s \in [t_{m-1}, t_m). \tag{3.1.12}$$

ii). For all n sufficiently large, and $m = 1, ..., F$,

$$\lambda^n_{v_m}\left(\phi_s\right) \geq c_1 \left(\prod_{j \in \mathcal{N}_m} \left(\phi_s\right)_j\right)^{p_1}, \quad s \in [t_{m-1}, t_m). \tag{3.1.13}$$

iii). $k \leq cn \left\|x - y\right\|$.

In the weakly interacting particle systems we consider, $\{\lambda_v\left(x\right)\}$ are specified by

(2.3.6) and (2.3.3), and each $\lambda_v(x)$ is therefore a linear combination of monomials in x multiplied by a Lipschitz function in x. Also, $\{\lambda_v^n(x)\}$ can be specified by (2.2.6) and (2.2.4). Thus the polynomial type lower bound in (3.1.10) and (3.1.13) holds for the particle systems.

Condition 3.1.14. *i). There exist constants $c', F, p, p_1 < \infty$ and $c, c_1 > 0$, such that for any $x, y \in \mathcal{S}$, there exists a strongly communicating path ϕ that connects x to y (with respect to c', F, c, p, c_1, p_1).*

ii). For any $n \in \mathbb{N}$ and any $x, y \in \mathcal{S}_n$, there exists a discrete strongly communicating path ϕ_n that connects x to y (with respect to c', F, c_1, p_1).

Clearly, Condition 3.1.14 is a strengthening of Condition 3.1.4. We now state the locally uniform LDP result, which is proved in Section 3.7.

Theorem 3.1.15. *Suppose the family of jump rates $\{\lambda_v^n(x), x \in \mathcal{S}, v \in \mathcal{V}\}_{n\in\mathbb{N}}$ satisfies Conditions 2.3.1, 3.1.1, 3.1.3 and 3.1.14. Also, assume the initial conditions $\{\mu^n(0)\}_{n\in\mathbb{N}}$ are deterministic, and $\mu^n(0) \to \mu_0 \in P(\mathcal{X})$ as n tends to infinity. Let $\{x_n\}_{n\in\mathbb{N}} \subset \mathcal{S}_n$, $x \in \mathcal{S}$ be such that $x_n \to x$ as $n \to \infty$. Then for any $t \in [0, 1)$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\mu^n(t) = x_n) = -J_t(\mu_0, x),$$

where J_t is as defined in (3.1.9).

3.1.3 LDP for invariant measures

In [20], a uniform sample path LDP for small noise diffusions (with respect to initial conditions) is used to study its metastability properties, including the mean exit time, most likely exit location from a given domain, and the LDP for invariant measures. The program of [20] was carried out for non-degenerate diffusions in $\mathbb{R}^d$; here we have a sequence of jump processes on lattice approximations of a compact set. However, we remark here that when Condition 3.1.14.i) holds, the same arguments carry through without essential change. In [20] extra conditions are assumed to guarantee that the process does not escape to infinity with significant probability; for our model, since the state space is compact, this is automatic. The quasipotential is defined by

$$V(x,y) = \inf\left\{I_t(\gamma) : \gamma \in D([0,t]:\mathcal{S}), \gamma(0) = x, \gamma(t) = y, t < \infty\right\}, \text{ for } x, y \in \mathcal{S}. \tag{3.1.14}$$

We need the quasipotential to be continuous within its domain. This property follows from the fact that for any $x, y \in \mathcal{S}$ that are sufficiently close, one can construct a path connects x to y with arbitrary small cost (Lemma 3.7.1).

Lemma 3.1.16. *Assume Condition 3.1.14.i) holds, then $V(\cdot,\cdot)$ is jointly continuous in $\mathcal{S} \times \mathcal{S}$.*

The proof of Lemma 3.1.16 is deferred to Section 3.7. We now state the LDP for invariant measures in the case of a single equilibrium.

Theorem 3.1.17. *Assume that x_0 is the unique stable equilibrium of the LLN dynamics (2.3.2), and is globally attracting (in $\mathcal{S}$). Also assume Condition 3.1.14.i) holds. Then for any $n > 0$, there exists a unique invariant measure ν^n of the Markov process with generator (2.2.5). Moreover, $\{\nu^n\}_{n\in\mathbb{N}}$ satisfies an LDP with rate function $V(x_0,\cdot)$.*

When the LLN limit (2.3.2) has multiple stable equilibria, following the same approach as the case of non-degenerate diffusions (see [20], Chapter 6.4), a generalization of the above theorem also holds.

3.1.4 Corollaries for the mean field interacting particle systems

We now describe how the results in Sections 3.1.1 and 3.1.2 apply to the particular setting of the mean field interacting particle systems introduced in Chapter 2.

Notice that in the statements of Theorems 3.1.10 and 3.1.15, one can rewrite the local rate function (3.1.3) as a function of the jump rates $\{\alpha^k_{\mathbf{ij}}\}$ (see (2.3.3)) of the particle systems. For this, we need to introduce a natural ordering for pairs $(\mathbf{i},\mathbf{j}) \in \cup_{k\le K}\mathcal{J}^k$, by first ordering them by increasing k, and then ordering $((i_1,..,i_k),(j_1,...,j_k))$ lexicographically. Denote by $\mathcal{C}_K$ the cardinality of $\cup_{k\le K}\mathcal{J}^k$. This ordering establishes a bijection between natural numbers $\{1,...,\mathcal{C}_K\}$ and ordered pairs $(\mathbf{i},\mathbf{j}) \in \cup_{k\le K}\mathcal{J}^k$. With this, we sometimes abuse the notation and refer to $r \in \mathbb{N}$ instead of $r = e_{\mathbf{j}} - e_{\mathbf{i}} = \sum_{l=1}^k e_{i_l} - \sum_{l=1}^k e_{j_l}$.

Let W be a $d \times \mathcal{C}_K$ matrix with each column specified as the corresponding jump direction: $W_r = e_{\mathbf{j}} - e_{\mathbf{i}}$ if $r \in \mathbb{N}$ maps to $e_{\mathbf{j}} - e_{\mathbf{i}}$. For $x \in \mathcal{S}$ and $\beta \in \Delta^{d-1}$, (3.1.3) can be written as

$$L(x,\beta) = \inf_{q:Wq=\beta} \sum_{k=1}^{K} \sum_{(\mathbf{i},\mathbf{j})\in\mathcal{J}^k} \alpha^k_{\mathbf{ij}}(x)\,\ell\left(\frac{q^k_{\mathbf{ij}}}{\alpha^k_{\mathbf{ij}}(x)}\right),$$

where $q = \operatorname{diag}\big(q^1,...,q^K\big)$ is a block matrix such that each q^k is a $d^k \times d^k$ matrix with nonnegative entries for $(\mathbf{i},\mathbf{j}) \in \mathcal{J}^k$, and zero for $(\mathbf{i},\mathbf{j}) \in \mathcal{X}^{2k}\backslash\mathcal{J}^k$. In the special case that $K=1$, L can be written as

$$L\left(x,\beta\right)=\inf_{q:\mathbf{1}^{T}q=\beta}\sum_{i=1}^{d}\sum_{j=1,j\neq i}^{d}\alpha_{ij}^{1}\left(x\right)\ell\left(\frac{q_{ij}}{\alpha_{ij}^{1}\left(x\right)}\right),$$

and the constraint $Wq=\beta$ reduces to

$$\sum_{i\neq j}q_{ij}-\sum_{k\neq j}q_{jk}=\beta_{j},\qquad j=1,...,d,$$

where we represent q^{1} simply by q.

Conditions 2.3.1, 3.1.1 and 3.1.7 are properties of $\{\lambda_{v}\left(x\right)\}$, and therefore can be directly verified from (2.3.6). Conditions 3.1.4 and 3.1.14 hold for many classes of interacting particle systems with simultaneous jumps. Here, we state three easily verifiable sufficient conditions in terms of the Γ matrices of the original particle systems. For $k\leq K$ and $(\mathbf{i},\mathbf{j})\in\mathcal{J}^{k}$, denote

$$\mathfrak{M}_{\mathbf{ij}}^{k}\doteq\inf_{x\in\mathcal{S}}\Gamma_{\mathbf{ij}}^{k}\left(x\right),\tag{3.1.15}$$

and let

$$\mathcal{N}^{k}\doteq\left\{(\mathbf{i},\mathbf{j})\in\mathcal{J}^{k}:\mathfrak{M}_{\mathbf{ij}}^{k}>0\right\}.$$

The first condition states that an entry of the rate matrices in a given jump direction is either identically zero, or uniformly bounded below away from zero (note that the jump rate $\alpha_{\mathbf{ij}}$ in that direction is not bounded away from zero).

Condition 3.1.18. *For $k=1,...,K$ and $(\mathbf{i},\mathbf{j})\in\mathcal{J}^{k}$, either $(\mathbf{i},\mathbf{j})\in\mathcal{N}^{k}$, or $\Gamma_{\mathbf{ij}}^{k}\left(x\right)=0$ for every $x\in\mathcal{S}$.*

Condition 3.1.19. *For every $x\in\mathcal{S}$, the Markov process on $\mathcal{X}$ with rate matrix $\Gamma^{1}\left(x\right)$ is ergodic.*

Condition 3.1.20. *For every $x \in \mathcal{S}$, the Markov process on $\mathcal{X}$ with rate matrix $\Gamma^{\text{eff}}(x)$ is ergodic, where Γ^{eff} is defined as in (2.3.5).*

In Section 3.2, we establish the following result.

Proposition 3.1.21. *The following two assertions hold:*

i). Conditions 3.1.18 and 3.1.20 imply Condition 3.1.4.

ii). Condition 3.1.19 and the continuity of $\Gamma^1(\cdot)$ implies Condition 3.1.14 (and hence, also Condition 3.1.4).

In particular, both the LDP and the locally uniform LDP hold for the interacting particle systems described in Examples 2.4.1, 2.4.2 and 2.4.3.

For the $n-$particle systems we study, it is more natural to start with random initial conditions. Depending on the large deviation rate of the initial condition, this gives rise to an additional cost in the rate function. The LDP for empirical measure processes with random initial conditions are stated in the following corollary.

Corollary 3.1.22. *Assume Conditions 2.3.1, 3.1.1, 3.1.3, 3.1.18 and 3.1.20 hold. Also, assume that the empirical measure processes $\{\mu^n\}_{n\in\mathbb{N}}$ are Markovian, with generator (2.2.3) and initial conditions $\{\mu^n(0)\}_{n\in\mathbb{N}}$ that converges to μ_0, in such a way that they satisfy an LDP with rate function $J_0(\cdot)$. Then the corresponding sequence of empirical measure processes $\{\mu^n\}_{n\in\mathbb{N}}$ satisfies the sample path LDP with rate function $J_0(\gamma(0)) + I(\gamma)$.*

Proof. For any $y \in \mathcal{S}_n$, denote $U^n(y) = -\frac{1}{n}\log \mathbb{E}_y e^{-nh(\mu^n)}$. By Theorem 3.1.10 and the equivalence between the LDP and the Laplace principle [40], for any sequence of

$\{y_n\}_{n\in\mathbb{N}}$ that converges to some $y \in \mathcal{S}$,

$$U^n\left(y_n\right) \to U\left(y\right) = \inf\left\{I\left(\gamma\right) + h\left(\gamma\right) : \gamma \in D\left([0,1] : \mathcal{S}\right), \gamma\left(0\right) = y\right\}.$$

By Lemma 3.1.23 below, U is continuous on $\mathcal{S}$. A standard argument by contradiction then shows that U^n converges to U uniformly. Let ν^n denote the law of $\mu^n\left(0\right)$. Then

$$\begin{aligned}
\lim_{n\to\infty} -\frac{1}{n}\log \mathbb{E}_{\mu^n(0)} e^{-nh(\mu^n)} &= \lim_{n\to\infty} -\frac{1}{n}\log \sum_{y_n\in\mathcal{S}_n} e^{-nU^n(y_n)}\nu^n\left\{y_n\right\} \\
&= \lim_{n\to\infty} -\frac{1}{n}\log \int e^{-n(U(y)+o(1))}\nu^n\left(dy\right) \\
&= \inf_{y\in\mathcal{S}}\left\{U\left(y\right) + J_0\left(y\right)\right\} \\
&= \inf_{\gamma\in D([0,1]:\mathcal{S})}\left\{h\left(\gamma\right) + J_0\left(\gamma\left(0\right)\right) + I\left(\gamma\right)\right\},
\end{aligned}$$

where the third equality follows from the assumption and the continuity of U. The conclusion follows by the equivalence between the LDP and the Laplace principle. □

Lemma 3.1.23. $U\left(y\right) = \inf\left\{I\left(\gamma\right) + h\left(\gamma\right) : \gamma \in D\left([0,1] : \mathcal{S}\right), \gamma\left(0\right) = y\right\}$ *is continuous.*

Proof. Fix $\varepsilon > 0$. Given $\delta > 0$ such that $c\left(\delta\right) \leq \varepsilon/3$, and any $y^\delta \in \mathcal{S}$ such that $\left\|y^\delta - y\right\| \leq \delta$, by Lemma 3.7.1, there exists a path $\nu \in AC\left([0,\delta] : \mathcal{S}\right)$ with $\nu\left(0\right) = y$, $\nu\left(\delta\right) = y^\delta$ such that $I_\delta\left(\nu\right) \leq c\left(\delta\right)$. And, by the construction in the proof of Lemma 3.7.1, $\sup_{s\in[0,\delta]}\left\|\nu\left(s\right) - y\right\| \leq C\delta$ for some $C < \infty$. We now rescale γ to obtain a new path γ^δ: for $c = \left(1-\delta\right)^{-1}$, define $\gamma^\delta \in AC\left([0,1-\delta] : \mathcal{S}\right)$ by $\gamma^\delta\left(s\right) = \gamma\left(cs\right)$. By Proposition 3.5.8, we can take δ smaller if necessary such that $I_{1-\delta}\left(\gamma^\delta\right) \leq I\left(\gamma\right) + \varepsilon/3$. Let $\bar{\gamma}$ be the concatenation of ν and γ^δ, $\left\|\gamma - \bar{\gamma}\right\|_\infty \to 0$ as $\delta \to 0$. Therefore

$$U\left(y^\delta\right) \leq h(\bar{\gamma}) + I\left(\bar{\gamma}\right) \leq I\left(\gamma\right) + \varepsilon/3 + c\left(\delta\right) + h\left(\bar{\gamma}\right) - h\left(\gamma\right),$$

which goes to zero by sending $\delta \to 0$ and then sending $\varepsilon \to 0$. The other inequality is proved in the same way. $\square$

Example 3.1.24. *Assume that in the $n-$particle system, particles are initially distributed as i.i.d $\mathcal{X}-$valued random variables, with common distribution ν. Then by Sanov's theorem,* $J_0(\mu_0) = R(\mu_0 \| \nu) = \sum_{i=1}^{d} \mu_{0,i} \log \frac{\mu_{0,i}}{\nu_i}$.

The analogous result holds for the locally uniform refinement of the LDP.

We note that these are only sufficient conditions, and the LDP can be shown to hold for other interacting particle systems, by directly verifying the conditions of Theorem 3.1.10 (especially, Condition 3.1.4), as illustrated by the following example.

Example 3.1.25. *Let $d = 4$, $K = 2$, and define the generator of the Markov process $\{\mu^n\}$ as in (2.2.3) with $\alpha_{\mathbf{ij}}^{k,n}$ defined as in (2.2.4), in terms of $\Gamma^{1,n}$ and $\Gamma^{2,n}$ given by*

$$\begin{aligned}
&\Gamma_{12}^{1,n}(x) = c_1, \quad \Gamma_{21}^{1,n}(x) = c_2, \quad && \Gamma_{34}^{1,n}(x) = c_3, \\
&\Gamma_{43}^{1,n}(x) = c_4, \quad \Gamma_{(1,2),(3,4)}^{2,n}(x) = c_5, \quad && \Gamma_{(3,4),(1,2)}^{2,n}(x) = c_6
\end{aligned}$$

for $c_i > 0$, $i = 1, ..., 6$, for all $x \in \mathcal{S}$, and identically zero otherwise. The effective jump rate matrix Γ^{eff} then takes the form

$$\begin{aligned}
&\Gamma_{12}^{\text{eff}}(x) = c_1, && \Gamma_{21}^{\text{eff}}(x) = c_2, && \Gamma_{34}^{\text{eff}}(x) = c_3, && \Gamma_{43}^{\text{eff}}(x) = c_4 \\
&\Gamma_{13}^{\text{eff}}(x) = x_2 c_5, && \Gamma_{31}^{\text{eff}}(x) = x_4 c_6, && \Gamma_{24}^{\text{eff}}(x) = x_1 c_5, && \Gamma_{42}^{\text{eff}}(x) = x_3 c_6.
\end{aligned}$$

Γ^{eff} is not ergodic on the part of the boundary given by
$\{x \in \mathcal{S} : x_3 = x_4 = 0 \text{ or } x_1 = x_2 = 0\}$. However, one can easily verify Condition 3.1.4, as well as Conditions 2.3.1, 3.1.1 and 3.1.7.

Given a mean field interacting particle system with simultaneous jumps, one can construct another single jump particle system, according to the effective jump rate matrix $\left\{\Gamma^{n,\mathrm{eff}}\right\}$ defined as in (2.3.4). As mentioned in Chapter 2, these systems have the same LLN limit. However, the following example shows that the prelimit models, in particular their large deviation behavior, can have a significantly different nature.

Example 3.1.26. *Let $\mathcal{X} = \{1,2,3,4,5,6\}$, $K = 2$, and for any $n \in \mathbb{N}$ consider a Markov process with the following properties: states are grouped into blocks $\{1,2\}$, $\{3,4\}$, $\{5,6\}$, such that within each block, states communicate through single jumps, and there exist simultaneous jumps of two particles between different blocks. Also, the number of particles in the first block can only change by an even number during each jump. In particular, consider the Markov process associated with the generator (2.2.3), with $\alpha^{k,n}_{\mathbf{ij}}$ defined as in (2.2.4), in terms of $\Gamma^{1,n}$ and $\Gamma^{2,n}$ given by*

$$\begin{aligned}
\Gamma^{1,n}_{12}(x) &= 1, \Gamma^{1,n}_{21}(x) = 1, \Gamma^{1,n}_{34}(x) = 1, \Gamma^{1,n}_{43}(x) = 1, \Gamma^{1,n}_{56}(x) = 1, \Gamma^{1,n}_{65}(x) = 1,\\
\Gamma^{2,n}_{(1,2),(3,4)}(x) &= 1, \quad \Gamma^{2,n}_{(3,4),(1,2)}(x) = 1, \quad \Gamma^{2,n}_{(2,3),(4,1)}(x) = 1, \quad \Gamma^{2,n}_{(4,1),(2,3)}(x) = 1,\\
\Gamma^{2,n}_{(1,3),(4,2)}(x) &= 1, \quad \Gamma^{2,n}_{(1,3),(5,2)}(x) = 1, \quad \Gamma^{2,n}_{(1,4),(5,2)}(x) = 1, \quad \Gamma^{2,n}_{(1,6),(5,2)}(x) = 1,\\
\Gamma^{2,n}_{(2,4),(1,5)}(x) &= 1, \quad \Gamma^{2,n}_{(2,5),(1,4)}(x) = 1, \quad \Gamma^{2,n}_{(1,5),(6,2)}(x) = 1, \quad \Gamma^{2,n}_{(3,5),(6,1)}(x) = 1,\\
\Gamma^{2,n}_{(4,6),(1,2)}(x) &= 1, \quad \Gamma^{2,n}_{(1,2),(5,6)}(x) = 1, \quad \Gamma^{2,n}_{(5,6),(1,2)}(x) = 1, \quad \Gamma^{2,n}_{(5,6),(3,4)}(x) = 1.
\end{aligned}$$

for all $x \in \mathcal{S}$, and identically zero otherwise. One computes the effective jump rate *matrix to be*

$$
\begin{array}{lll}
\Gamma^{\text{eff}}_{12}(x) = 1, & \Gamma^{\text{eff}}_{21}(x) = 1 + x_4 + x_5, & \Gamma^{\text{eff}}_{34}(x) = 1, \\
\Gamma^{\text{eff}}_{43}(x) = 1, & \Gamma^{\text{eff}}_{56}(x) = 1, & \Gamma^{\text{eff}}_{65}(x) = 1, \\
\Gamma^{\text{eff}}_{13}(x) = x_2 + x_4, & \Gamma^{\text{eff}}_{24}(x) = x_1 + x_3, & \Gamma^{\text{eff}}_{31}(x) = x_2 + x_4, \\
\Gamma^{\text{eff}}_{42}(x) = 2x_1 + x_3, & \Gamma^{\text{eff}}_{14}(x) = x_3, & \Gamma^{\text{eff}}_{32}(x) = 2x_1, \\
\Gamma^{\text{eff}}_{15}(x) = x_2 + x_3 + x_4 + x_6, & \Gamma^{\text{eff}}_{62}(x) = x_1 + x_4 + x_5, & \Gamma^{\text{eff}}_{45}(x) = x_2, \\
\Gamma^{\text{eff}}_{54}(x) = x_2, & \Gamma^{\text{eff}}_{16}(x) = x_5, & \Gamma^{\text{eff}}_{52}(x) = x_1, \\
\Gamma^{\text{eff}}_{36}(x) = x_5, & \Gamma^{\text{eff}}_{51}(x) = x_3 + x_6, & \Gamma^{\text{eff}}_{41}(x) = x_6, \\
\Gamma^{\text{eff}}_{53}(x) = x_6, & \Gamma^{\text{eff}}_{64}(x) = x_5, & \Gamma^{\text{eff}}_{26}(x) = x_1.
\end{array}
$$

Conditions 2.3.1, 3.1.1, 3.1.3 and 3.1.18 can be checked explicitly from (2.3.3) and (2.3.6). Also, Condition 3.1.20 follows from the fact that $\Gamma^{\text{eff}}(x)$ *is affine in* x*, and* $\{\Gamma^{\text{eff}}(e_i)\}$ *is ergodic for* $i \in \{1, ..., 6\}$*. Thus, the associated single jump particle system satisfies an LDP and a locally uniform LDP by Theorem 3.1.10 and 3.1.15. In particular, for any* $x, y \in \mathcal{S}$*, if we take* $x^n, y^n \in \mathcal{S}^n$ *such that* $\|x^n - x\| \to 0$*,* $\|y^n - y\| \to 0$*, Theorem 3.1.15 implies*

$$
\mathbb{P}_{x_n}(\mu^n(t) = y_n) = \exp(-nJ_t(x, y) + o(n)), \text{ as } n \to \infty,
$$

where $J_t(\cdot)$ *is defined as in (3.1.9) and is finite because of the communication property. However, the original simultaneous jump particle system fails to satisfy the locally uniform LDP. Indeed, for any* $n \in \mathbb{N}$*, the number of particles in states* $\{1, 2\}$ *remains either odd or even. Therefore if we take* $x^n, y^n \in \mathcal{S}^n$ *that additionally satisfy* $\{x_1^n + x_2^n\}_{n\in\mathbb{N}}$ *is even and* $\{y_1^n + y_2^n\}_{n\in\mathbb{N}}$ *is odd, then* $\mathbb{P}_{x_n}(\mu^n(t) = y_n) = 0$ *for any* $n \in \mathbb{N}$*.*

3.2 Verification of Conditions

In this section we verify that the assumptions of Theorems 3.1.10 and 3.1.15 are satisfied by a large class of mean field interacting particle systems. Readers only interested in the large deviation proof can skip this section.

Recall the definition of $\mathfrak{M}^k_{\mathbf{ij}}$ in (3.1.15), and let

$$c_0 = \min\left\{\mathfrak{M}^k_{\mathbf{ij}} : \mathfrak{M}^k_{\mathbf{ij}} > 0, k = 1, ..., K, (\mathbf{i}, \mathbf{j}) \in \mathcal{J}^k\right\} > 0.$$

3.2.1 Proof of Proposition 3.1.21.i)

To prove Condition 3.1.4 under Conditions 3.1.18 and 3.1.20, we will actually prove a stronger result. Namely, we show that Condition 3.1.4 holds under a much weaker communication condition for the Markov processes on $\mathcal{X}$, introduced below as Definition 3.2.1.

Definition 3.2.1. *For two states $u, v \in \mathcal{X}$, v is said to be $K-$**accessible** from u, if there exist $L \in \{1, ..., d\}$ and a sequence of distinct states in $\mathcal{X}$: $u = u_1, u_2, ..., u_L = v$, such that for $m = 1, ..., L-1$,*

i). there exist $k_m \in \{1, ..., K\}$, $(\mathbf{i}_m, \mathbf{j}_m) \in \mathcal{J}^{k_m}$, and $l_m, l_m' \in \{1, ..., k_m\}$, such that $u_m = i_{m,l_m}$ and $u_{m+1} = j_{m,l_m'}$,

ii). for $l = 1, ..., k_m$, $i_{m,l} \in \{u_1, ..., u_m\}$,

iii). $\mathfrak{M}^{k_m}_{\mathbf{i}_m\mathbf{j}_m} > 0$.

*If for any $u, v \in \mathcal{X}$, v is $K-$accessible from u, we call the Markov process associated with $\{\Gamma_{\mathbf{ij}}(\cdot)\}_{(\mathbf{i},\mathbf{j})\in\cup_{k\leq K}\mathcal{J}^k}$ $K-$**ergodic.***

Notice that when $K = 1$, the notion of $1-$ergodicity coincides with the ergodicity of single jump rate matrix Γ^1. Roughly speaking, the general $K-$accessibility condition ensures that one can reach state v from u, by going through certain types of simultaneous jumps, such that for each m, one can move mass exclusively from a subset of states $\{u_1, ..., u_m\}$ to u_{m+1} (this is ensured by the second condition of Definition 3.2.1). Up to our knowledge, all the interesting examples satisfy the conditions of Definition 3.2.1. And, the second condition allows one to construct communication paths, by moving along the direction of such sequences of simultaneous jumps.

To illustrate the notion of $K-$ergodicity, note that the interacting particle system with simultaneous jumps described in Examples 3.1.25 and 3.1.26 are $2-$ergodic. Let us verify $2-$ergodicity for Example 3.1.25. Take any $i, j \in \{1, 2, 3, 4\}$, and assume without loss of generality $i = 1$, by the symmetry of the dynamics. If $j = 2$, we can take $L = 2$, $k_1 = 1$ and $(\mathbf{i}_1, \mathbf{j}_1) = (1, 2)$. If $j = 3$, take $L = 3$, $u_1 = 1$, $u_2 = 2$, $u_3 = 3$, $k_1 = 1$, $(\mathbf{i}_1, \mathbf{j}_1) = (1, 2)$, $k_2 = 2$ and $(\mathbf{i}_2, \mathbf{j}_2) = ((1, 2), (3, 4))$. The $j = 4$ case is similar to $j = 3$, except that $u_3 = 4$. It is easy to check the sequence of states $\{u_m\}$ and jumps satisfy conditions i), ii), iii) of Definition 3.2.1.

Remark 3.2.2. *We show that Conditions 3.1.18 and 3.1.20 imply $K-$ergodicity of $\{\Gamma_{\mathbf{ij}}(x)\}_{(\mathbf{i},\mathbf{j})\in\cup_{k\leq K}\mathcal{J}^k}$, for any $x \in \mathcal{S}$. Take any $u, v \in \mathcal{X}$, $u \neq v$. Since $\Gamma^{\text{eff}}(e_u)$ is ergodic by Condition 3.1.20, there exists a sequence of distinct states $u = u_1, ..., u_L = v$ such that $\Gamma^{\text{eff}}_{u_m u_{m+1}}(e_u) > 0$, $m = 1, ..., L-1$. By the definition of Γ^{eff} given in (2.3.5), this implies that for $m = 1, ..., L-1$, there exist $k_m \in \{1, ..., K\}$, $(\mathbf{i}_m, \mathbf{j}_m) \in$*

$\mathcal{J}^{k_m}$, *and* $l_m \in \{1, ..., k_m\}$, *such that* $u_m = i_{m,l_m}$, $u_{m+1} = j_{m,l_m}$ *and*

$$\prod_{\substack{r=1 \\ r \neq l_m}}^{k_m} \left\langle e_u, e_{i_{m,r}} \right\rangle \Gamma^{k_m}_{\mathbf{i}_m \mathbf{j}_m} (e_u) > 0.$$

This implies that $\mathfrak{M}^{k_m}_{\mathbf{i}_m \mathbf{j}_m} > 0$, *and* $\mathbf{i}_{m,r} = u$ *for every* $r \neq l_m$. *In other words, the* l_m^{th} *component of* $\mathbf{i}_m$ *is equal to* u_m, *and all other components are equal to* u. *Definition 3.2.1.i) is satisfied with* $l_m = l_m^{'}$, *and Definition 3.2.1.ii) is satisfied with* $i_{m,l} \in \{u, u_m\}$ *for* $l = 1, ..., k_m$. *Finally, since* $\Gamma^{k_m}_{\mathbf{i}_m \mathbf{j}_m} (e_u) > 0$, *Condition 3.1.18 implies that* $\mathfrak{M}^{k_m}_{\mathbf{i}_m \mathbf{j}_m} > 0$, *property iii) of Definition 3.2.1 is also satisfied.*

To prove Condition 3.1.4, one needs to show some local controllability property of the underlying Markov process. More precisely, it suffices to show that there exist constants $c > 0$ and $c', p < \infty$ such that for every $x \in \mathcal{S}$ and $y \in \text{int}(\mathcal{S})$, there exists a communicating path ϕ from x to y (with respect to the constants c, p and c'), as stated in Definition 3.1.4. We show that $K-$ergodicity provides sufficient controllability. First, it allows one to move from any point on the boundary to some compact convex subset of $\text{int}(\mathcal{S})$ that contains y, using a sequence of jumps whose rates are uniformly bounded below away from zero (Lemma 3.2.3). Then, we show that in the compact subset, the controllability is stronger: one can in fact move toward any coordinate direction (Lemma 3.2.4), and thus, the construction of such a path in this compact subset is straightforward (Lemma 3.2.6).

For $a > 0$ define $\tilde{\mathcal{S}}^a \doteq \{x \in \mathcal{S} : x_i \geq a, \, i = 1, ..., d\}$.

Lemma 3.2.3. *Assume that for any* $i, j \in \mathcal{X}$, j *is* $K-$*accessible from* i. *Then for any* $x \in \mathcal{S}$ *and* $a \in [0, 1/(K+1)^{d-1} d]$, *there exist* $z \in \tilde{\mathcal{S}}^a$, $t_0 > 0$ *and a communicating path* $\phi \in C([0, t_0] : \mathcal{S})$ *that connects* x *to* z. *Moreover,* *i). there is $C < \infty$ such that for any $x \in \mathcal{S}$, ϕ can be chosen such that $\int_0^{t_0} \|\dot{\phi}(s)\| ds \leq C\, dist(x, \tilde{\mathcal{S}}^a)$,*

ii). for any $s \in [0, t_0]$ and $i \in \mathcal{X}$ such that $\dot{\phi}_i(s) < 0$, $\phi_i(s) \geq a$.

Before proving the lemma in general let us first illustrate with Example 3.1.25 where $d = 4$ and $K = 2$. Assume without loss of generality $x_1 \geq x_2 \geq x_3 \geq x_4$, and therefore $x_1 \geq 1/4$. As before, we take $(\mathbf{i}_1, \mathbf{j}_1) = (1, 2)$, and $(\mathbf{i}_2, \mathbf{j}_2) = ((1, 2), (3, 4))$. For $a \leq 1/108$, we discuss the construction of such a path in three cases that exhaust all possibilities.

Case I. $x_3 \geq a > x_4$. Take $\phi(0) = x$, $\dot{\phi}(t) = (e_{\mathbf{j}_1} - e_{\mathbf{i}_1}) 1_{[0, 2(a - x_4))}(t)$ $+ (e_{\mathbf{j}_2} - e_{\mathbf{i}_2}) 1_{[2(a-x_4), 3(a-x_4))}(t)$. It is easy to verify property i) in the lemma. To verify property ii), note that for $t \in [0, 3(a - x_4)]$, $\phi_1(t) \geq \phi_1(0) - 2(a - x_4) - (a - x_4) \geq x_1 - 3a \geq a$, $\phi_2(t) \geq \phi_2(0) \geq a$, $\dot{\phi}_3(t) \geq 0$, $\dot{\phi}_4(t) \geq 0$. Since $\phi_4(3(a - x_4)) = a$, we have $\phi(3(a - x_4)) \in \tilde{\mathcal{S}}^a$.

Case II. $x_2 \geq a > x_3$. Take $\phi^{(1)}(0) = x$, $\dot{\phi}^{(1)}(t) = (e_{\mathbf{j}_1} - e_{\mathbf{i}_1}) 1_{[0, 2(a - x_3))}(t) + (e_{\mathbf{j}_2} - e_{\mathbf{i}_2}) 1_{[2(a-x_3), 3(a-x_3))}(t)$. One can verify property i) and ii) in the lemma are satisfied for $\phi^{(1)}$ by the same argument applied to ϕ. If $z^{(1)} = \dot{\phi}^{(1)}(3(a - x_3))$, then we have $z_1^{(1)} \geq a$, $z_2^{(1)} \geq a$, $z_3^{(1)} \geq a$. Then it is reduced to Case I, and we can further take a communicating path that moves $z^{(1)}$ into $\tilde{\mathcal{S}}^a$, and concatenates with $\phi^{(1)}$ to obtain the desired path.

Case III. $x_1 > a > x_2$. Take $\phi^{(2)}(0) = x$, $\dot{\phi}^{(2)}(t) = (e_{\mathbf{j}_1} - e_{\mathbf{i}_1}) 1_{[0, a - x_2)}(t)$. One can verify property i) and ii) in the lemma are satisfied for $\phi^{(2)}$. And at $t = a - x_2$, $x_1 \geq a = x_2 \geq x_3 \geq x_4$, and it reduces to Case II.

The construction of a communication path can be generalized into the following proof.

Proof. If $a = 0$ or $x \in \tilde{\mathcal{S}}^a$, we can choose $z = x$ and there is nothing to prove. Therefore we assume $a > 0$ and $x \notin \tilde{\mathcal{S}}^a$. Assume without loss of generality that $x_1 \geq x_2 \geq \cdots \geq x_d$, and for some $2 \leq k \leq d$, $x_{k-1} \geq a > x_k$. When $K = 1$, a more straightforward proof (which gives stronger result) is given in Lemma 3.2.7 below. For general K, we will prove the lemma by induction on $w \doteq d - k$, one less than the number of indicies smaller than a.

If $w = 0$, then x_d is the only component such that $x_d < a$. By the definition of $K-$accessibility there exist $L \in \{1, ..., d\}$ and a sequence of distinct states $1 = u_1, u_2, ..., u_L = d$, such that for $m = 1, ..., L-1$, there exist $k_m \in \{1, ..., K\}$, $(\mathbf{i}_m, \mathbf{j}_m) \in \mathcal{J}^{k_m}$, and $l_m, l_m' \in \{1, ..., k_m\}$, such that $u_m = i_{m,l_m}$, $u_{m+1} = j_{m,l_m'}$, and $\mathfrak{M}^{k_m}_{\mathbf{i}_m \mathbf{j}_m} > 0$. Now let

$$c_{m,L} \doteq \begin{cases} (K+1)^{L-2} - (K+1)^{L-1-m} & \text{if } m \in \{1, ..., L-1\}, \\ (K+1)^{L-2} & \text{if } m = L. \end{cases}$$

Then consider the piecewise linear path ϕ with $\phi(0) = x$, and

$$\dot{\phi}(t) = \sum_{m=1}^{L-1} (e_{\mathbf{j}_m} - e_{\mathbf{i}_m}) 1_{[c_{m,L}(a-x_d), c_{m+1,L}(a-x_d))}(t), \text{ a.e. } t \geq 0. \tag{3.2.1}$$

Let $t_0 = (K+1)^{L-2}(a - x_d)$. We claim that for $l = 2, ..., d-1$, and $t^* \in [0, t_0]$ such that $\dot{\phi}_l(t^*) < 0$, $\phi_l(s) \geq a$ for $s \in [t^*, t_0]$. Indeed, for such l, there exist $m \in \{1, ..., L-1\}$ and $r \in \{1, ..., k_m\}$ such that $l = i_{m,r}$. If there exist multiple (m, r) that satisfy $l = i_{m,r}$, we take the pair with smallest m. Then by Definition 3.2.1.ii), $l \in \{u_1, ..., u_{m-1}\}$, thus there exist $m' \in \{1, ..., m-1\}$ and $r' \in \{1, ..., k_{m'}\}$

such that $l = j_{m',r'}$. Therefore, for any $s \in [t^*, t_0]$,

$$\begin{aligned}
\phi_l(s) &= \phi_l(0) + \int_0^s \dot{\phi}_l(t)\,dt \\
&\geq a + \left\langle e_{\mathbf{j}_{m'}}, e_l \right\rangle (c_{m'+1,L} - c_{m',L})(a - x_d) \\
&- \sum_{p=m}^{L-1} \left\langle e_{\mathbf{i}_p}, e_l \right\rangle (c_{p+1,L} - c_{p,L})(a - x_d) \\
&\geq a,
\end{aligned}$$

where the last inequality follows from the fact that $\left\langle e_{\mathbf{i}_p}, e_l \right\rangle \leq K$, and

$$c_{m'+1,L} - c_{m',L} \geq c_{m,L} - c_{m-1,L} \geq K \sum_{p=m}^{L-1} (c_{p+1,L} - c_{p,L}).$$

Also, as will explained below, for $s \in [0, t_0]$,

$$\begin{aligned}
\phi_1(s) = \phi_1(0) &+ \int_0^s \dot{\phi}_1(t)\,dt \\
&\geq \frac{1}{d} - \sum_{p=1}^{L-1} \left\langle e_{\mathbf{i}_p}, e_1 \right\rangle (c_{p+1,L} - c_{p,L})(a - x_d) \\
&\geq \frac{1}{d} - K c_{L,L} \frac{1}{(K+1)^{d-1} d} \\
&\geq \frac{1}{d} - \frac{1}{(K+1)\,d} \\
&> a.
\end{aligned}$$

Here the first inequality follows since $\sum_{i=1}^{d} x_i = 1$ and $x_1 \geq \cdots \geq x_d$ implies $x_1 \geq 1/d$. The second inequality follows from $\left\langle e_{\mathbf{i}_p}, e_1 \right\rangle \leq K$ and $a - x_d \leq a \leq \frac{1}{(K+1)^{d-1} d}$. The third inequality follows since $c_{L,L} = (K+1)^{L-2} \leq (K+1)^{d-2}$, and the fourth since $K/K+1 \geq 1/(K+1)^{d-1}$.

Finally, by Definition 3.2.1.ii), $\dot{\phi}_d(s) \geq 0$, and since $\langle e_{\mathbf{j}_{L-1}} - e_{\mathbf{i}_{L-1}}, e_d \rangle \geq 1$,

$$\begin{aligned}
\phi_d(t_0) &= x_d + \int_0^{t_0} \dot{\phi}_d(t)\,dt \\
&\geq x_d + \int_{c_{L-1,L}(a-x_d)}^{c_{L,L}(a-x_d)} \dot{\phi}_d(t)\,dt \\
&\geq x_d + \langle e_{\mathbf{j}_{L-1}} - e_{\mathbf{i}_{L-1}}, e_d \rangle (c_{L,L} - c_{L-1,L})(a - x_d) \\
&\geq x_d + a - x_d = a.
\end{aligned}$$

Therefore $\phi(t_0) \in \tilde{\mathcal{S}}^a$, and property ii) of Lemma 3.2.3 is satisfied. Also, ϕ is a communicating path, since by (2.3.3) and (2.3.6),

$$\begin{aligned}
\lambda_{v_m}(\phi(s)) &\geq \alpha_{\mathbf{i}_m\mathbf{j}_m}^{k_m}(\phi(s)) \\
&= \frac{1}{(k_m)!}\prod_{r=1}^{k_m} \phi_{i_{m,r}}(s)\,\Gamma_{\mathbf{i}_m\mathbf{j}_m}^{k_m}(\phi(s)) \\
&\geq \frac{1}{K!}a^d c_0.
\end{aligned}$$

Property i) is satisfied since $\int_0^{t_0} \|\dot{\phi}(s)\|ds \leq \sqrt{2K}t_0 \leq C_2(d,K)(a - x_d)$ $\leq C_2(d,K)\mathrm{dist}(x, \tilde{\mathcal{S}}^a)$ for some $C_2 < \infty$.

Suppose the conclusion of the lemma holds for $w \leq h$, and now let $w = h+1$ (and $k = d - h - 1$). The idea is first move along a communicating path that uses the same jump directions as (3.2.1), until reaching some $z^{(1)}$, such that for some $s \in \{k, ..., d\}$, $z_s^{(1)} \geq a$, and then use the induction assumption. To construct such a path $\phi^{(1)}$, take the same sequence of jumps $\{(\mathbf{i}_m, \mathbf{j}_m)\}$ as in the case of $w = 0$. Let m_0 be the smallest index such that one of the states with value below a can be increased: $m_0 = \min\{m : \exists l \text{ s.t. } \exists s \in \{k, ..., d\}, s = j_{m,l}\}$. Let

$\Delta = \min\{a - x_s : s \in \{k, ..., d\}, \text{ s.t. } \exists l,\ s = j_{m_0,l}\}$. We then define

$$c_{m,m_0} = \begin{cases} (K+1)^{m_0-2} - (K+1)^{m_0-1-m} & m \in \{1, ..., m_0 - 1\}, \\ (K+1)^{m_0-2} & m = m_0, \end{cases}$$

and let $\phi^{(1)}$ be the path given by $\phi^{(1)}(0) = x$,

$$\dot{\phi}^{(1)}(t) = \sum_{m=1}^{m_0-1} \left(e_{\mathbf{j}_m} - e_{\mathbf{i}_m}\right) 1_{[c_{m,m_0}\Delta, c_{m+1,m_0}\Delta)}(t), \text{ a.e. } t \geq 0.$$

Also, let $z^{(1)} = \phi^{(1)}(c_{m_0,m_0}\Delta)$. Then by the same argument applied to the path ϕ, $\phi^{(1)}$ is a communicating path, and for $t^* \in [0, c_{m_0,m_0}\Delta]$ and $l \in \mathcal{X}$ such that $\dot{\phi}_l^{(1)}(t^*) < 0$, it follows that $\phi_l^{(1)}(s) \geq a$ for $s \in [t^*, c_{m_0,m_0}\Delta]$. Also, $\int_0^{c_{m_0,m_0}\Delta} \|\dot{\phi}^{(1)}(s)\| ds \leq \sqrt{2K} c_{m_0,m_0}\Delta \leq C_2(d,K)\text{dist}(x, \tilde{\mathcal{S}}^a)$. Therefore, for $l \in \{1, ..., k-1\}$, $z_l^{(1)} \geq a$, and there exists $s \in \{k, ..., d\}$, such that $z_s^{(1)} \geq a$. Thus at most h components of $z^{(1)}$ are less than a. By the induction assumption there exists $z \in \tilde{\mathcal{S}}^a$, and a communicating path $\phi^{(2)}$ that connects $z^{(1)}$ to z, and satisfies properties i) and ii). The conclusion then follows by concatenating $\phi^{(1)}$ and $\phi^{(2)}$, and noticing that

$$\begin{aligned} \text{dist}\left(z^{(1)}, \tilde{\mathcal{S}}^a\right) &\leq \|x - z^{(1)}\| + \text{dist}\left(x, \tilde{\mathcal{S}}^a\right) \\ &\leq \int_0^{c_{m_0,m_0}\Delta} \|\dot{\phi}^{(1)}(s)\| ds + \text{dist}\left(x, \tilde{\mathcal{S}}^a\right) \\ &\leq (C_2(d,K) + 1)\,\text{dist}\left(x, \tilde{\mathcal{S}}^a\right). \end{aligned}$$

□

Lemma 3.2.4. *Assume that for any $i, j \in \mathcal{X}$, j is $K-$accessible from i. Then for any $u, v \in \mathcal{X}$, $u \neq v$, there exist a finite constant $F = F_{u,v}$, $k_m \in \{1, ..., K\}$ for*

$m = 1, ..., F$, $a_m \geq 0$ and $(\mathbf{i}_m, \mathbf{j}_m) \in \mathcal{J}^{k_m}$ such that $\mathfrak{M}^{k_m}_{\mathbf{i}_m \mathbf{j}_m} > 0$, such that

$$e_v - e_u = \sum_{m=1}^{F} a_m \left(e_{\mathbf{j}_m} - e_{\mathbf{i}_m}\right). \tag{3.2.2}$$

We first illustrate with Example 3.1.25. Without loss of generality, we set $v = 4$ and $u = 1$. As before, we take $(\mathbf{i}_1, \mathbf{j}_1) = (1, 2)$, and $(\mathbf{i}_2, \mathbf{j}_2) = ((1, 2), (3, 4))$. Thus $(e_{\mathbf{j}_1} - e_{\mathbf{i}_1}) + (e_{\mathbf{j}_2} - e_{\mathbf{i}_2}) = e_3 + e_4 - 2e_1$.

To cancel the term e_3, we further take $(\mathbf{i}_3, \mathbf{j}_3) = (3, 4)$. Then $\sum_{m=1}^{3} (e_{\mathbf{j}_m} - e_{\mathbf{i}_m}) = 2(e_4 - e_1)$, or $e_4 - e_1 = \sum_{m=1}^{3} \frac{1}{2} (e_{\mathbf{j}_m} - e_{\mathbf{i}_m})$.

Proof of Lemma 3.2.4. Fix $u, v \in \mathcal{X}$, $u \neq v$. When $K = 1$, we can take a sequence of states $u = u_1, ..., u_L = v$, such that $\mathfrak{M}^1_{u_m u_{m+1}} > 0$. Then we can simply take $e_v - e_u = \sum_{m=1}^{L-1} \left(e_{u_{m+1}} - e_{u_m}\right)$.

For general K, it is more subtle to choose all the coefficients $\{a_m\}$. By Definition 3.2.1, there exist a sequence of distinct states $u = u_1, ..., u_L = v$, and $k_m \in \{1, ..., K\}$, $(\mathbf{i}_m, \mathbf{j}_m) \in \mathcal{J}^{k_m}$ for $m = 1, ..., L-1$, that satisfies the properties in Definition 3.2.1. We claim that there exist nonnegative $\{a^{(u)}_m\}_{m=1}^{L-1}$, such that

$$\sum_{m=1}^{L-1} a^{(u)}_m \left(e_{\mathbf{j}_m} - e_{\mathbf{i}_m}\right) = \sum_{i \neq u} c^{(u)}_i e_i - e_u \tag{3.2.3}$$

with $c^{(u)}_i \geq 0$, $\sum_{i \neq u} c^{(u)}_i = 1$, $c^{(u)}_v > 0$.

To prove this claim, we first let $\kappa^{(m)} = \left\langle e_{u_{m+1}}, e_{\mathbf{j}_m} \right\rangle / k_m$ for $m = 1, ..., L-1$. Since by Definition 3.2.1 $u_{m+1} \in \mathbf{j}_m$, $\kappa^{(m)} \in [1/K, 1]$. Let $\{b^{(u)}_m\}_{m=1}^{L-1}$ be the solution to the following system of linear equations:

$$\begin{cases} b_{L-1}^{(u)} = 1 \\ \kappa^{(L-2)} b_{L-2}^{(u)} = K b_{L-1}^{(u)} \\ \dots \\ \kappa^{(1)} b_1^{(u)} = K \left(b_{L-1}^{(u)} + b_{L-2}^{(u)} + \dots + b_2^{(u)} \right). \end{cases} \tag{3.2.4}$$

The interpretation of these equations is as follows. Recall that by Definition 3.2.1.ii), for each $m = 1, \dots, L$, each component of $\mathbf{i}_m$ belongs to $\{u_1, \dots, u_m\}$. Hence for $m_0 = 2, \dots, L-1$, $\langle e_{u_{m_0}}, e_{\mathbf{i}_m} \rangle$ can be positive only when $m \in \{m_0, \dots, L-1\}$. In the following display, the first equality follows from the last sentence, the second equality from the definition of $\kappa^{(m_0-1)}$, and the inequality from $\langle e_{\mathbf{i}_m}, e_{u_{m_0}} \rangle \leq K$ and the equations (3.2.4):

$$\begin{aligned} \left\langle b_{m_0-1}^{(u)} e_{\mathbf{j}_{m_0}-1} - \sum_{m=1}^{L-1} b_m^{(u)} e_{\mathbf{i}_m}, e_{u_{m_0}} \right\rangle &= \left\langle b_{m_0-1}^{(u)} e_{\mathbf{j}_{m_0-1}} - \sum_{m=m_0}^{L-1} b_m^{(u)} e_{\mathbf{i}_m}, e_{u_{m_0}} \right\rangle \\ &= \kappa^{(m_0-1)} k_{m_0-1} b_{m_0-1}^{(u)} - \sum_{m=m_0}^{L-1} b_m^{(u)} \langle e_{\mathbf{i}_m}, e_{u_{m_0}} \rangle \\ &\geq 0. \end{aligned} \tag{3.2.5}$$

Equality holds in (3.2.5) only if $k_{m_0-1} = 1$ and if for all $m \in \{m_0, \dots, L-1\}$, $\langle e_{\mathbf{i}_m}, e_{u_{m_0}} \rangle = K$. Since equality for a particular $m_0 \in \{2, \dots, L-1\}$ implies $k_{m_0-1} = 1$ and therefore $\langle e_{\mathbf{i}_{m_0-1}}, e_{u_{m_0-1}} \rangle < K$, there must be some index for which the strict inequality holds. It follows from (3.2.5) that

$$\left\langle \sum_{m=1}^{L-1} b_m^{(u)} (e_{\mathbf{j}_m} - e_{\mathbf{i}_m}), e_i \right\rangle \geq \left\langle b_{m_0-1}^{(u)} e_{\mathbf{j}_{m_0}-1} - \sum_{m=1}^{L-1} b_m^{(u)} e_{\mathbf{i}_m}, e_i \right\rangle \geq 0, \text{ for } i \in \{u_m\}_{m=2}^{L-1}.$$

Also, for $i \notin \{u_m\}_{m=1}^{L-1}$, by Definition 3.2.1.ii), $\langle e_i, e_{\mathbf{i}_m} \rangle = 0$ for any $m = 1, \dots, L$. This implies

$$\left\langle \sum_{m=1}^{L-1} b_m^{(u)} \left(e_{\mathbf{j}_m} - e_{\mathbf{i}_m}\right), e_i \right\rangle = \left\langle \sum_{m=1}^{L-1} b_m^{(u)} e_{\mathbf{j}_m}, e_i \right\rangle \geq 0, \text{ for } i \notin \{u_m\}_{m=1}^{L-1}.$$

The equality in the following display is because $\langle e_{\mathbf{j}_m} - e_{\mathbf{i}_m}, (1, \ldots, 1)\rangle = 0$, and the inequality follows by combining the last two displays and using that the first will be strict for at least one index:

$$\left\langle \sum_{m=1}^{L-1} b_m^{(u)} \left(e_{\mathbf{j}_m} - e_{\mathbf{i}_m}\right), e_u \right\rangle = -\sum_{w \neq u} \left\langle \sum_{m=1}^{L-1} b_m^{(u)} \left(e_{\mathbf{j}_m} - e_{\mathbf{i}_m}\right), e_w \right\rangle < 0.$$

The last three displays imply there are $d_i^{(u)} \geq 0$, $i \neq u$ and $d^* > 0$ such that

$$\sum_{m=1}^{L-1} b_m^{(u)} \left(e_{\mathbf{j}_m} - e_{\mathbf{i}_m}\right) = \sum_{i \neq u} d_i^{(u)} e_i - d^* e_u.$$

To obtain (3.2.3) and in particular the coefficients $\{a_m^{(u)}\}_{m=1}^{L-1}$ and $\{c_i^{(u)}\}_{i=1, i\neq u}^{d}$, we divide the last display by d^*. Then clearly $c_i^{(u)} \geq 0$ for $i \neq u$, and again using $\langle e_{\mathbf{j}_m} - e_{\mathbf{i}_m}, (1, \ldots, 1)\rangle = 0$ gives $\sum_{i \neq u} c_i^{(u)} = 1$. To see that $c_v^{(u)} > 0$, we use the last display and that $v \notin \mathbf{i}_m$ for any m, $v \in \mathbf{j}_{L-1}$, and that $b_m^{(u)} > 0$ for all m.

To obtain (3.2.2) we will eliminate the terms involving e_i, $i \neq u, v$, on the right hand side of (3.2.3). We have assumed for each $s \neq u, v$, that v is $K-$accessible from s. Hence applying the same argument as above with u replaced by s, we obtain some $L_s < \infty$, a sequence of jumps $\{(\mathbf{i}_m^{(s)}, \mathbf{j}_m^{(s)})\}_{m=1}^{L_s-1}$ and coefficient $\{a_m^{(s)}\}_{m=1}^{L_s-1}$, such that

$$\sum_{m=1}^{L_s-1} a_m^{(s)} \left(e_{\mathbf{j}_m^{(s)}} - e_{\mathbf{i}_m^{(s)}}\right) = \sum_{i \neq s} c_i^{(s)} e_i - e_s, \tag{3.2.6}$$

where $c_i^{(s)} \geq 0$, $\sum_{i \neq s} c_i^{(s)} = 1$, $c_v^{(s)} > 0$, and $a_m^{(s)} \geq 0$.

It suffices to find nonnegative $\{\theta_i\}_{i=1,i\neq v}^d$ such that

$$\theta_1\left(\sum_{i\neq 1} c_i^{(1)} e_i - e_1\right) + \cdots + \theta_d\left(\sum_{i\neq d} c_i^{(d)} e_i - e_d\right) = e_v - e_u, \tag{3.2.7}$$

since then we coul substitute (3.2.6) into (3.2.7) to obtain (3.2.2). Notice that (3.2.7) can also be written componentwiseas

$$\begin{cases} \theta_1 = c_1^{(2)}\theta_2 + \cdots + c_1^{(d)}\theta_d \\ \theta_2 = c_2^{(1)}\theta_1 + \cdots + c_2^{(d)}\theta_d \\ \vdots \\ \theta_u = 1 + c_u^{(1)}\theta_1 + \cdots + c_u^{(d)}\theta_d \\ \vdots \\ \theta_d = c_d^{(1)}\theta_1 + \cdots + c_d^{(d-1)}\theta_{d-1}. \end{cases}$$

Applying Lemma 3.2.5 below with

$$C = \begin{pmatrix} 0 & c_1^{(2)} & \dots & c_1^{(d)} \\ c_2^{(1)} & \ddots & & \vdots \\ \vdots & & \ddots & c_{d-1}^{(d)} \\ c_d^{(1)} & \dots & c_d^{(d-1)} & 0 \end{pmatrix}$$

and $b = e_u$, it follows that this system of linear equations has a unique nonnegative solution. □

Lemma 3.2.5. *Suppose that $A = I_d - C$, where I_d is the $d \times d$ identity matrix and $C = (c_{ij})_{i,j=1}^d$ for some $d \in \mathbb{N}$, such that $c_{ii} = 0$, $c_{ij} \geq 0$ and $\sum_{j=1}^d c_{ij} < 1$. Also, let $b \in [0,\infty)^d$. Then the system of linear equations $Ax = b$ has a unique nonnegative solution $\{x_i\}_{i=1}^d$.*

Proof. The spectral radius of C is less than 1 since its matrix norm is less than 1.

Therefore $\det A > 0$, and A^{-1} exists. The fact that A^{-1} is a positive matrix follows from a general result in inverse positivity [2, Theorem 6.3.8]. The nonnegativity of x then follows from the nonnegativity of b. $\square$

Lemma 3.2.6. *Assume the conclusion of Lemma 3.2.4 holds. Then for any $a > 0$ and $x, y \in \mathcal{S}^a$, there exists a communicating path in $\mathcal{S}^{a/2}$ that connects x to y, with the number of linear segments possibly depending on a.*

Proof. To start we use the fact that there is a piecewise linear path ϕ_0 from x to y that lies in $\mathcal{S}^a$ and only uses velocities in the directions $\{e_j - e_i : i, j \in \mathcal{X}\}$, and for which the total length scales with the distance between x and y. Suppose ϕ_0 is such a path, with $\phi_0(0) = x$, and for some $L < d$, there are $0 = t_0 < t_1 < \cdots < t_L$ and $\{u_m\}_{m=1}^L, \{v_m\}_{m=1}^L \in \mathcal{X}$ so that

$$\dot{\phi}_0(t) = \sum_{m=1}^{L} (e_{v_m} - e_{u_m}) 1_{[t_{m-1}, t_m)}(t), \text{ a.e. } t \in [0, t_L].$$

As noted previously, we can assume there is $C_1 < \infty$ independent of x and y such that $\int_0^{t_L} \|\dot{\phi}_0(s)\| ds \leq C_1 \|x - y\|$. By Lemma 3.2.4, there exist $F_m < \infty$, $\{(\mathbf{i}_k^{(m)}, \mathbf{j}_k^{(m)})\}$ and $\{a_k^{(m)}\}$ such that $e_{v_m} - e_{u_m} = \sum_{k=1}^{F_m} a_k^{(m)} (e_{\mathbf{j}_k^{(m)}} - e_{\mathbf{i}_k^{(m)}})$. Let $F = \max_m F_m$, $C = \sqrt{2K} F \max_{m,k} a_k^{(m)}$, and set $\varepsilon = a/2C$. We now further divide each segment $[t_{m-1}, t_m)$ into finitely many segments, each of which has length no more than ε, and within each segment, replace ϕ_0 by a finitely segmented communicating path. In particular, if $[r_i, r_f) \subset [t_{m-1}, t_m)$ is such a segment, then we would use the path

$$\phi(s) = \phi_0(r_i) + \int_{r_i}^{s} \sum_{k=1}^{F_m} a_k^{(m)} F_m \left(e_{\mathbf{j}_k^{(m)}} - e_{\mathbf{i}_k^{(m)}} \right) 1_{[r_i + \frac{k-1}{F_m}(r_f - r_i), r_i + \frac{k}{F_m}(r_f - r_i))}(s)\, ds.$$

Note that $\phi(r_f) = \phi_0(r_i)$, and that $r_f - r_i \leq \varepsilon$ for all such segments implies $\phi \in AC\left([0, t_L] : \mathcal{S}^{a/2}\right)$. Also, by (2.3.3) and (2.3.6) we have $\lambda_v(\phi(s)) \geq \frac{1}{K!} \left(\frac{a}{2}\right)^d c_0$ for all

$s \in [0, t_L]$, and

$$\int_0^{t_L} \|\dot{\phi}(s)\| ds \le CF^2 \int_0^{t_L} \|\dot{\phi}_0(s)\| ds \le CF^2 C_1 \|x-y\|.$$

Thus ϕ is the desired communicating path that connects x to y. $\square$

Now we complete the proof of Proposition 3.1.21.i).

Proof. By Remark 3.2.2, under Conditions 3.1.18 and 3.1.20 j is $K-$accessible from i for any $i, j \in \mathcal{X}$. Therefore for any $x \in \mathcal{S}$, $y \in \text{int}(\mathcal{S})$, a communicating path connecting x to y can be found by setting $a = \min\left\{y_1, ..., y_d, \frac{1}{(K+1)^{d-1} d}\right\}$, and concatenating the paths constructed in Lemma 3.2.3 and 3.2.4. Notice that along the path ϕ, $\lambda_v(\phi(s)) \ge \frac{1}{K!}\left(\frac{a}{2}\right)^d c_0 \ge c_1 (\min_i y_i)^d$, where $c_1 = \frac{1}{K!} c_0 (2(K+1)^{d-1} d)^{-d}$. Therefore Condition 3.1.4 holds. $\square$

3.2.2 Proof of Proposition 3.1.21.ii)

To verify Condition 3.1.14, for any $x, y \in \mathcal{S}$, we need to construct a strongly communicating path that connects x to y. In particular, we need to consider the case when $y \in \partial\mathcal{S}$. Thus, the approach in the previous subsection (in particular, Lemma 3.2.6) cannot be directly applied. We will show that $1-$ergodicity implies a very strong controllability, so that one can always construct a communicating path between x and y by matching up their coordinates. We first show that Condition 3.1.14 holds under both Conditions 3.1.18 and 3.1.19, and then show that 3.1.18 can be dropped if Γ^1 is continuous.

Lemma 3.2.7. *Assume Conditions 3.1.18 and 3.1.19 hold. Then for any* $x, y \in$

$\mathcal{S}$, there exist $t_0 > 0$ and a strongly communicating path $\phi \in AC([0,t_0]:\mathcal{S})$ that connects x to y. Moreover, we can rescale ϕ such that $\phi(0) = x$, $\phi(1) = y$, and $\|\dot{\phi}(\cdot)\|$ is constant and bounded by $C\|x-y\|$ for some $C < \infty$.

Proof. We will prove the result by a recursive construction. We claim that for any $r \in \{2,...,d\}$, after possibly relabeling indicies, there exist $0 \le t_{r-1} < \infty$, some $z^{(r)} \in \mathcal{S}$ such that $z_i^{(r)} = y_i$ for $i \in \{1,...,r-1\}$, and a strongly communicating path $\phi^{(r)} \in AC([0,t_{r-1}]:\mathcal{S})$, such that $\phi^{(r)}(0) = x$, $\phi^{(r)}(t_{r-1}) = z^{(r)}$. Furthermore, $\int_0^{t_{r-1}} \|\dot{\phi}^{(r)}(s)\|ds \le C_r(d,K)\|x-y\|$ for some $C_r < \infty$. The conclusion follows by taking $r = d$.

The claim will be proved by induction. The case $r = 2$ is trivial: we choose $i_1 \in \mathcal{X}$ so that $x_{i_1} \ge y_{i_1}$, and then reverse the roles of the indicies 1 and i_1. Take $t_1 = x_1 - y_1$ and any state $i_2 \in \mathcal{X}\backslash\{1\}$ such that $\mathfrak{M}^1_{1i_2} > 0$ (the existence of i_2 is implied by Conditions 3.1.18 and 3.1.19), and switch the indicies 2 and i_2. Define

$$\phi^{(2)}(t) = x + (e_2 - e_1)t,\ t \in [0,t_1].$$

Then clearly, $\phi^{(2)}(0) = x$, $z^{(2)} = \phi^{(2)}(t_1) \in \mathcal{S}$, $z_1^{(2)} = y_1$ and $\int_0^{t_1}\|\dot{\phi}^{(2)}(s)\|ds = \sqrt{2}t_1 \le \sqrt{2}\|x-y\|$. Thus, $\phi^{(2)}$ is a path of the desired form. Also, the lower bound (3.1.13) holds with $\mathcal{N}_1 = \{1\}$, $p_1 = 1$, and for some $c_1 > 0$ since by (2.3.3) and (2.3.6),

$$\lambda_{v_1}(x) = \sum_{k=1}^{K} \sum_{\substack{(\mathbf{i},\mathbf{j})\in\mathcal{J}^k: \\ e_{\mathbf{j}}-e_{\mathbf{i}}=v_1}} \alpha^k_{\mathbf{ij}}(x) \ge x_1\Gamma^1_{12}(x) \ge c_0 x_1$$

since $\Gamma^1_{12}(x)$ is uniformly positive.

Now, assume the claim holds for $r = m$, and let $z^{(m)}$ and $\phi^{(m)}$ be the corresponding quantities in the claim. We now prove the claim for $r = m+1$. By assumption, we have $\phi^{(m)}(t_{m-1}) = z^{(m)}$. Since $\sum_{i=m}^{d} z_i^{(m)} = \sum_{i=m}^{d} y_i$, by possibly indicies within the set $\{m, ..., d\}$, we may assume without loss of generality that $z_m^{(m)} \geq y_m$. If $z_m^{(m)} = y_m$ then the claim also holds for $r = m+1$. If $z_m^{(m)} > y_m$, we will move some mass from state m to some state in $\{m+1, ..., d\}$, while not changing the mass in states with lower index, by considering an additional trajectory $\psi \in AC([0,q] : \mathcal{S})$ for some $q < \infty$, such that $\psi_i(q) = \psi_i(0)$ for every $i = 1, ..., m-1$ and $\psi_m(q) - \psi_m(0) = -(z_m^{(m)} - y_m)$. To do this, take any $v \in \{m+1, ..., d\}$. By Conditions 3.1.18 and 3.1.19, there exist $L \leq m$ and a sequence of states $m = u_0, ..., u_L = v$, such that for $1 \leq l \leq L-1$, $\mathfrak{M}^1_{u_l u_{l+1}} > 0$. We take $i_m = u_{l_{\min}}$, where $l_{\min} \doteq \min\{l : u_l \in \{m+1, ..., d\}\}$. Thus $u_{l_{\min}}$ is the first index outside $\{1, ..., m\}$, and $l_{\min}$ is the number of steps it took to get there. Define $q = l_{\min}(z_m^{(m)} - y_m)$, and

$$\psi(t) = z^{(m)} + \int_0^t \sum_{l=1}^{l_{\min}} \left(e_{u_l} - e_{u_{l-1}}\right) 1_{[(l-1)(z_m^{(m)} - y_m), l(z_m^{(m)} - y_m))}(s)\, ds. \tag{3.2.8}$$

Since $l_{\min} \leq d$, we have

$$\begin{aligned}
\int_0^{t_p} \|\dot{\phi}^p(s)\| ds &\leq \sqrt{2}d \left|z_m^{(m)} - y_m\right| \\
&\leq \sqrt{2}d \left(|x_m - y_m| + \left|z_m^{(m)} - x_m\right|\right) \\
&\leq \sqrt{2}d \left(\|x - y\| + \int_0^{t_{m-1}} \|\dot{\phi}^{(m)}(s)\| ds\right) \\
&\leq \sqrt{2}d \left(1 + C_m(d, K)\right) \|x - y\|,
\end{aligned}$$

where the last inequality follows from the induction assumption for $r = m$.

Now define $t_{m+1} = t_m + q$, and $\phi^{(m+1)} \in C([0, t_{m+1}] : \mathcal{S})$ to be the concatenation of $\phi^{(m)}$ and ψ. As we show below, $\phi^{(m+1)}$ is a path of the desired form. For $v \in \mathcal{V}_1$, $v = e_j - e_i$ for some $i, j \in \mathcal{X}$, $i \neq j$, as before we have $\lambda_v(x) \geq x_i \Gamma^1_{ij}(x) \geq c_0 x_i$, and therefore for $s \in [0, t_{m+1}]$, $\lambda_v(\phi^{(m+1)}(s)) \geq c_0 \phi_i^{(m+1)}(s)$. Finally we note that

$$\begin{aligned}
\int_0^{t_m} \|\dot{\phi}^{(m+1)}(s)\| ds &\leq \int_0^{t_{m-1}} \|\dot{\phi}^{(m)}(s)\| ds + \int_0^{t_p} \|\dot{\psi}(t)\| dt \\
&\leq \left(\sqrt{2}d + \left(\sqrt{2}d + 1\right) C_m(d, K)\right) \|x - y\|.
\end{aligned}$$

The claim now follows by induction. The last sentence in the statement of the lemma follows from Remark 3.1.6, thus completing the proof. □

The construction of a discrete strongly communicating path follows directly from discretization of a strongly communicating path.

Lemma 3.2.8. *Suppose Conditions 3.1.18 and 3.1.19 hold. Then for any $n \in \mathbb{N}$, $x, y \in \mathcal{S}_n$, there exists a discrete strongly communicating path that connects x to y.*

Proof. Fix $n \in \mathbb{N}$ and $x, y \in \mathcal{S}_n$. Let $y - x = \frac{1}{n}(a_1, .., a_d)$ with $a_i \in \mathbb{N}$, $|a_i| \leq n \|x - y\|$ and $\sum_{i=1}^d a_i = 0$. By Lemma 3.2.7 there exists a communicating path $\phi^{(c)}$ with $\phi^{(c)}(0) = x$, $\phi^{(c)}(1) = y$, and $U < \infty$ such that $\|\dot{\phi}^{(c)}(s)\| \doteq U \leq C(d, K)\|x - y\|$ for almost every $s \in [0, 1]$. Since $\phi^{(c)}$ only moves in the directions $\{e_j - e_i : i, j \in \mathcal{X}\}$, the choice of the times q used in the recursive construction of $\phi^{(c)}$ in (3.2.8) and the fact that both x and y lie in the grid $\mathcal{S}_n$ imply that the value $\phi^{(c)}$ at the times when $\dot{\phi}^{(c)}$ changes (denoted $\{z^{(r)}\}$ in the previous lemma) are also in $\mathcal{S}_n$. It suffices to take $k = nU$, and the discrete path $\{\phi_s\}_{s=1}^k$ to be $\phi_s = \phi^{(c)}(s/nU)$, that is, to be the lattice point passed by $\phi^{(c)}$. Then ϕ is a path of the desired form and ϕ satisfies (3.1.13) since $\phi^{(c)}$ satisfies (3.1.10). □

We now relax the assumptions made in the last two lemmas, and show that they continue to hold under Condition 3.1.19 and continuity of the jump rates. The basic idea of the proof is to partition $\mathcal{S}$ into finite collection of sets of essentially the same geometric form as the simplex, such that within each sets Condition 3.1.18 holds. Therefore, by Lemma 3.2.7 and 3.2.8, one can construct a strongly communicating path (and its discrete analogue) that connects any two points within each small simplex. The desired strongly communicating path (and its discrete analogue) in $\mathcal{S}$ can be obtained by concatenating the paths within some collection of the smalll simplices.

Lemma 3.2.9. *Assume Condition 3.1.19 is satisfied, and that $\Gamma^1(\cdot)$ is continuous. Then the conclusions of Lemma 3.2.7 and 3.2.8 hold.*

Proof. Recall that $\mathcal{V}_1 \doteq \{e_j - e_i, (i,j) \in \mathcal{J}^1\}$. For any $x \in \mathcal{S}$, let $\mathcal{V}(x) = \{e_j - e_i \in \mathcal{V}_1 : \Gamma^1_{ij}(x) > 0\}$ and $c_x = \min_{v \in \mathcal{V}(x)} \Gamma^1_{ij}(x)$. By the continuity of $\{\Gamma^1_{ij}(\cdot)\}$, there exists an open simplex $\mathcal{S}_x$ that contains x, and such that $\inf_{y \in \mathcal{S}_x} \min_{e_j - e_i \in \mathcal{V}(x)} \Gamma^1_{ij}(y) \geq c_x/2$. $\{\mathcal{S}_x\}_{x \in \mathcal{S}}$ thus forms an open cover of $\mathcal{S}$, and we can take a finite subcover $\{\mathcal{S}_{x_k}\}_{k=1}^K$. Let $c_0 = \min\{c_{x_1}/2, ..., c_{x_K}/2\}$. For any $x \in \mathcal{S}$ let $K(x) \doteq \{k : x \in \mathcal{S}_{x_k}\}$, and define

$$\bar{\Gamma}^1_{ij}(x) = \begin{cases} \Gamma^1_{ij}(x) & e_j - e_i \in \cup_{k \in K(x)} \mathcal{V}(x_k) \\ 0 & e_j - e_i \in \mathcal{V} \setminus \cup_{k \in K(x)} \mathcal{V}(x_k). \end{cases}$$

Then for any $i = 1, ..., K$, $x \in \mathcal{S}_{x_i}$, the Markov process on $\mathcal{X}$ with rate matrices $\{\bar{\Gamma}^1_{ij}(x)\}$ is ergodic. This implies that $\{\bar{\Gamma}^1_{ij}(x)\}$ satisfies Conditions 3.1.18 and 3.1.19 within $\mathcal{S}_{x_k}$, with $\inf_{y \in \mathcal{S}_{x_k}} \min_{e_j - e_i \in \mathcal{V}_{x_k}} \bar{\Gamma}^1_{ij}(y) \geq c_0$. Therefore, by applying the same argument as in Lemma 3.2.7, it follows that for any $u, v \in \bar{\mathcal{S}}_{x_i}$, there exists a strongly communicating path that connects u to v.

Now for any $u, v \in \mathcal{S}$, suppose the line $\{tu + (1-t)v : t \in [0,1]\}$ intersects $\cup_{k=1}^K \partial \mathcal{S}_{x_i}$ at $\{u_i\}_{i=1}^L$, for some $L < \infty$. Then each line segment $\{tu_i + (1-t)u_{i+1} : t \in [0,1]\}$ is contained in some $\bar{\mathcal{S}}_{x_j}$, and therefore one can take a strongly communicating path ϕ_i that connects u_i to u_{i+1}. Let ϕ be the concatenation of $\{\phi_i\}_{i=0}^{L}$. Then ϕ is strongly communicating because each ϕ_i is. Indeed, the fact that each ϕ_i satisfies Condition 3.1.4.i).ii). and Definition 3.1.12.iv).v) directly implies ϕ does. Also, denote $u_0 = u$, $u_{L+1} = v$, we have

$$Len\left(\phi\right) \leq \sum_{i=0}^{L} Len\left(\phi_i\right) \leq C \sum_{i=0}^{L} \left\|u_{i+1} - u_i\right\| = C \left\|u - v\right\|,$$

for some $C < \infty$, where we used the fact that each ϕ_i satisfies Condition 3.1.4.iii). Therefore ϕ satisfies Condition 3.1.4.iii).

The construction of a discrete strongly communicating path follows by exactly the same argument as Lemma 3.2.8. $\square$

3.2.3 Proof of Lemma 3.1.9

As before, we first illustrate the conclusion with Example 3.1.25 where $d = 4$ and $K = 2$. Let $x = \mu(0)$, and assume without loss of generality $x_1 \geq x_2 \geq x_3 \geq x_4$, and therefore $x_1 \geq 1/4$. The ODE (2.3.2) implies

$$\dot{\mu}_1 = \sum_{v \in \mathcal{V}} \langle v, e_1 \rangle \lambda_v \left(\mu\left(t\right)\right) \geq -\mu_1 - \mu_1\mu_2 \geq -2\mu_1,$$

thus $\mu_1(t) \geq x_1 e^{-2t} \geq c_0$, $c_0 = e^{-2}/4$. Then

$$\dot{\mu}_2 = \sum_{v \in \mathcal{V}} \langle v, e_2 \rangle \lambda_v \left(\mu\left(t\right)\right) \geq \mu_1 - \mu_2 - \mu_1\mu_2 \geq c_0 - 2\mu_2,$$

which one solves $\mu_2(t) \geq c_1 t$ for some $c_1 > 0$. Also, for $i = 3, 4$,

$$\dot{\mu}_i = \sum_{v \in \mathcal{V}} \langle v, e_i \rangle \lambda_v(\mu(t)) \geq \mu_1 \mu_2 - \mu_i - \mu_3 \mu_4 \geq c_0 c_1 t - 2\mu_i,$$

thus $\mu_i(t) \geq c_2 t^2$. The proof for the general case is more technical and is given below.

Proof. Define $b_1 = \max\{|\langle e_j, v \rangle| : j \in \mathcal{X}, v \in \mathcal{V}\} < \infty$, and let $M_1 = b_1 M |\mathcal{V}|$. The ODE (2.3.2) implies that for $i = 1, ..., d$,

$$\dot{\mu}_i(t) = \sum_{v \in \mathcal{V}} \langle v, e_i \rangle \lambda_v(\mu(t)) \geq -M_1, \quad t \in [0, 1]. \tag{3.2.9}$$

Therefore

$$\mu_i(t) \geq \mu_i(0) e^{-M_1 t}, \quad t \in [0, 1]. \tag{3.2.10}$$

Let $\mathcal{X}_1 = \left\{ i \in \mathcal{X} : \mu_i(0) > \frac{1}{2d} \right\}$. Then for any $b < \frac{1}{2d} e^{-M_1}$, (3.2.10) implies that $\mu_i(t) > b$, $t \in [0, 1]$ holds for all $i \in \mathcal{X}_1$, thus Condition 3.1.3 is satisfied.

To show the inequality in Condition 3.1.3 also holds for $i \in \mathcal{X} \backslash \mathcal{X}_1$ (for a suitable choice of b), the idea is that in (3.2.9), for those v such that $\langle v, e_i \rangle < 0$, $\lambda_v(x)$ converges to zero as $x_i \to 0$ (by Condition 3.1.7), and therefore the communication condition would push $\mu(\cdot)$ into the interior of $\mathcal{S}$. Define $y \doteq \left(\frac{1}{d}, ..., \frac{1}{d} \right)$, and note that by Condition 3.1.4 and Remark 3.1.6, there exist $c_1 > 0$, and a communicating path $\phi \in AC([0, 1] : \mathcal{S})$ with $\phi(0) = \mu(0)$, $\phi(1) = y$. Given the representation of ϕ in terms of $F < \infty$, $\{t_m\}_{m=1}^F, \{v_m\}_{m=1}^F$ and $\{U_m\}_{m=1}^F$ given in (3.1.2), property ii) of Condition 3.1.4 implies for $m = 1, ..., F$,

$$\lambda_{v_m}(\phi(s)) \geq c_1, \; s \in [t_{m-1}, t_m]. \tag{3.2.11}$$

For $m=1,...,F$, denote $x^{(m)}=\phi(t_m)$. Then in particular, we have $\lambda_{v_{m+1}}\left(x^{(m)}\right)\geq c_1$. Using Condition 3.1.1, there exist $C_2, C_3>0$ such that the following lower bound holds for any $z\in\mathcal{S}$:

$$\frac{\lambda_{v_{m+1}}(z)}{\lambda_{v_{m+1}}\left(x^{(m)}\right)}\geq C_2\prod_{j=1}^{d}\frac{z_j}{x_j^{(m)}}\exp\left(-f_0\left(\left\|x^{(m)}-z\right\|\right)\right)\geq C_3\prod_{j=1}^{d}\frac{z_j}{x_j^{(m)}}, \tag{3.2.12}$$

where the last inequality follows from the fact that $f_0(\cdot)$ is continuous (and thus bounded from above) on $[0,1]$.

Now, define $\mathcal{P}_k=\{j\in\mathcal{X}:\mu_j(0)<x_j^{(k)}\}$, and fix $i\in\mathcal{X}\backslash\mathcal{X}_1$. Clearly, $i\in\mathcal{P}_F$ since $x_i^{(F)}=y_i=1/d$. We claim that for every $k=1,...,F$, there exists $b^{(k)}>0$ such that for any $j\in\mathcal{P}_k$, $\mu_j(t)\geq b^{(k)}t^{D(k)}$, where $D(k)=\sum_{i=0}^{k-1}d^i$. Since $i\in\mathcal{P}_F$, the lemma follows from setting $k=F$ and $b=b^{(F)}$ in the claim.

To prove the claim, we first note that by (3.2.10), for any $m=1,...,F$ and $j\notin\mathcal{P}_m$, $\mu_j(t)\geq e^{-M_1}\mu_j(0)\geq e^{-M_1}x_j^{(m)}$. For $t>0$, applying (3.2.12) to obtain

$$\frac{\lambda_{v_{m+1}}(\mu(t))}{\lambda_{v_{m+1}}\left(x^{(m)}\right)}\geq C_3e^{-M_1|\mathcal{P}_m^c|}\prod_{j\in\mathcal{P}_m}\frac{\mu_j(t)}{x_j^{(m)}}\geq C_3e^{-M_1d}\prod_{j\in\mathcal{P}_m}\mu_j(t). \tag{3.2.13}$$

Notice that if for some $j\in\mathcal{P}_m$, $x_j^{(m)}=0$, the second inequality in (3.2.13) also holds trivially.

We now prove the claim by induction. Define
$b_2=\min\{\langle v,e_j\rangle : j\in\mathcal{X},\ v\in\mathcal{V}, \text{ s.t. } \langle v,e_j\rangle>0\}$. For $k=1$, take any $j\in\mathcal{P}_1$. Since $\mu_j(0)<x_j^{(1)}$, we have $\langle v_1,e_j\rangle>0$. By (2.3.2), (3.2.11) and Condition 3.1.7,

$$\begin{aligned}\dot{\mu}_j(t) &\geq \langle v_1,e_j\rangle\lambda_{v_1}(\mu(t))+\sum_{\substack{v\in\mathcal{V}\\ \langle v,e_j\rangle<0}}\langle v,e_j\rangle\lambda_v(\mu(t))\\ &\geq b_2c_1-M_2\mu_j(t)\end{aligned}$$

for $M_2 = b_1 C |\mathcal{V}|$, where C is the constant in Condition 3.1.7. Applying the comparison principle for ODEs, we see that for $t \in [0,1]$, $\mu_j(t) \geq \frac{b_2 c_1}{M_2}\left(1 - e^{-M_2 t}\right) \geq b^{(1)} t$ for some $b^{(1)} > 0$.

Assume the claim holds for $k \leq m$, and let $k = m+1$. For $l \in \mathcal{P}_{m+1}$, since $\mu_l(0) < x_l^{(m+1)} = \phi_l(t_{m+1})$, and ϕ has representation (3.1.2), there exists $m^* \in \{1, ..., m+1\}$, such that $\langle v_{m^*}, e_l \rangle > 0$, and therefore $\langle v_{m^*}, e_l \rangle \geq b_2$. If $m^* = 1$, the result follows by applying the same argument in the $k = 1$ case. Note that by (3.2.13) and the induction hypothesis for $k = m$, there exist $C_4 > 0$ such that when $m^* \geq 2$, for $t \in [0,1]$,

$$\begin{aligned}
\frac{\lambda_{v_{m^*}}(\mu(t))}{\lambda_{v_{m^*}}\left(x^{(m^*-1)}\right)} &\geq C_3 e^{-M_1 d} \prod_{j \in \mathcal{P}_{m^*-1}} \mu_j(t) \geq C_3 e^{-M_1 d}\left(b^{(m^*-1)} t^{\sum_{i=0}^{m^*-2} d^i}\right)^d \\
&\geq C_4 t^{\sum_{i=1}^{m^*-1} d^i}.
\end{aligned}$$

Thus, combined with (2.3.2), (3.2.11) and Condition 3.1.7, there exist $C_5 > 0$ such that that

$$\begin{aligned}
\dot{\mu}_l(t) &\geq \langle v_{m^*}, e_l \rangle \lambda_{v_{m^*}}(\mu(t)) + \sum_{\substack{v \in \mathcal{V} \\ \langle v, e_l \rangle < 0}} \langle v, e_l \rangle \lambda_v(\mu(t)) \\
&\geq C_5 t^{\sum_{i=1}^{m^*-1} d^i} - M_2 \mu_l(t). \\
&\geq C_5 t^{\sum_{i=1}^{m} d^i} - M_2 \mu_l(t).
\end{aligned}$$

Solving the ODE $\dot{u}(t) = C_5 t^{\sum_{i=1}^{m} d^i} - M_2 u(t)$, and applying the comparison principle, it follows that for $t \in [0,1]$,

$$\begin{aligned} \mu_l(t) &\geq \frac{C_5}{M_2^{\sum_{i=0}^m d^i}} e^{-M_2 t} (-1)^{\sum_{i=0}^m d^i} \int_{-M_2 t}^0 s^{\sum_{i=1}^m d^i} e^{-s} ds \\ &\geq \frac{C_5}{M_2^{\sum_{i=0}^m d^i}} e^{-M_2 t} (-1)^{\sum_{i=0}^m d^i} \int_{-M_2 t}^0 s^{\sum_{i=1}^m d^i} ds \\ &\geq b^{(m+1)} t^{\sum_{i=0}^m d^i} \end{aligned}$$

for some $b^{(m+1)} > 0$. This proves the claim, and hence the lemma. □

3.3 The Variational Representation Formula

3.3.1 Variational Representation for a Poisson Random Measure

We briefly review the variational representation formula for a Poisson random measure stated in [6]. For any Polish space S let $M_\sigma(S)$ denote the space of $\sigma-$finite measures on S. We equip $M_\sigma(S)$ with the weakest topology such that for every $f \in C_c(S)$, the function $\nu \mapsto \int_S f d\nu, \nu \in M_\sigma(S)$, is continuous. Let $\mathcal{Y} = [0,\infty)$, $\mathcal{Y}_T = [0,T] \times \mathcal{Y}$, both equipped with the usual Euclidean topology, and let $\mathcal{M} = M_\sigma(\mathcal{Y}_T)$. For some fixed measure $\nu \in M_\sigma(\mathcal{Y})$, let $\nu_T = m_T \otimes \nu$, where m_T is Lebesgue measure on $[0,T]$. For $\theta \in [0,\infty)$, let $\mathbb{P}_\theta$ denote the unique probability measure on $(\mathcal{M}, \mathcal{B}(\mathcal{M}))$ under which the canonical map $N : \mathcal{M} \to \mathcal{M}$, $N(\omega) = \omega$, is a Poisson random measure with intensity measure $\theta\nu_T$. Let $\mathbb{E}_\theta$ denote expectation with respect to $\mathbb{P}_\theta$. For notational convenience, we omit the dependence of $\mathbb{P}_\theta$ and $\mathbb{E}_\theta$ on the fixed measure ν_T.

Now, define a controlled Poisson random measure as follows. Let $\mathcal{W} = \mathcal{Y} \times [0, \infty)$ and $\mathcal{W}_T = [0, T] \times \mathcal{W} = \mathcal{Y}_T \times [0, \infty)$, also equipped with the Euclidean product topology. Let $\bar{\mathcal{M}} = M_\sigma(\mathcal{W}_T)$ and let $\bar{\mathbb{P}}$ be the unique probability measure on $(\bar{\mathcal{M}}, \mathcal{B}(\bar{\mathcal{M}}))$ under which the canonical map $\bar{N} : \bar{\mathcal{M}} \to \bar{\mathcal{M}}$, $\bar{N}(\omega) = \omega$, is a Poisson random measure with intensity measure $\overline{\nu}_T = \nu_T \otimes m$, where m is Lebesgue measure on $[0, \infty)$. Let $\bar{\mathbb{E}}$ denote expectation with respect to $\bar{\mathbb{P}}$. Also, define

$$\mathcal{G}_t = \sigma \left\{ \bar{N}((0, s] \times A) : 0 \leq s \leq t, A \in \mathcal{B}(\mathcal{W}) \right\},$$

and let $\mathcal{F}_t$ denote its completion under $\bar{\mathbb{P}}$. Denote by $\bar{\mathcal{P}}$ the predictable σ-field on $[0, T] \times \bar{\mathcal{M}}$ with the filtration $\{\mathcal{F}_t\}_{0 \leq t \leq T}$ on $(\bar{\mathcal{M}}, \mathcal{B}(\bar{\mathcal{M}}))$.

Definition 3.3.1. *Let $\bar{\mathcal{A}}$ be the class of $(\bar{\mathcal{P}} \otimes \mathcal{B}(\mathcal{Y})) \backslash \mathcal{B}[0, \infty)$ predictable maps $\varphi : [0, T] \times \bar{\mathcal{M}} \times \mathcal{Y} \to [0, \infty)$.*

The role of φ is to control the intensity of jumps at (s, ω, x) by thinning in the r variable. For $\varphi \in \bar{\mathcal{A}}$, define $N^\varphi : \bar{\mathcal{M}} \to \mathcal{M}$ by

$$N_\omega^\varphi((0, t] \times U) = \int_{(0,t] \times U} \int_0^\infty 1_{[0, \varphi(s, \omega, x)]}(r) \, \bar{N}_\omega(dsdxdr), \tag{3.3.1}$$

for $t \in [0, T], U \in \mathcal{B}(\mathcal{Y}), \omega \in \bar{\mathcal{M}}$. We will suppress the dependence of $\varphi(t, \omega, x)$, $\bar{N}_\omega$ and N_ω^φ on ω at later times. Then under $\bar{\mathbb{P}}$, N^φ is a controlled random measure on $\mathcal{Y}_T$ with $\varphi(s, x)$ determining the intensity for points at location x and time s. With some abuse of notation, for $\theta \in [0, \infty)$ we will let N^θ be defined as in (3.3.1) with $\varphi(s, x) \equiv \theta$. Note that the law (on $\bar{\mathcal{M}}$) of N^θ under $\bar{\mathbb{P}}$ coincides with the law of N under $\mathbb{P}_\theta$.

Recall $\ell(\cdot)$ as defined in (3.1.4). For $\varphi \in \bar{\mathcal{A}}$ define the random variable $L_T(\varphi)$ by

$$\begin{aligned} L_T(\varphi)(\omega) &= \int_{\mathcal{Y}_T} \ell(\varphi(t,\omega,x))\, \nu_T(dtdx) \\ &= \int_0^T \left(\int_{[0,\infty)} \ell(\varphi(t,\omega,x))\, \nu(dx) \right) dt, \ \omega \in \bar{\mathcal{M}}. \end{aligned} \tag{3.3.2}$$

Definition 3.3.2. *Define $\bar{\mathcal{A}}_b$ to be the class of $\left(\bar{\mathcal{P}} \otimes \mathcal{B}(\mathcal{Y})\right) \backslash \mathcal{B}[0,\infty)$ predictable maps φ such that for some $B < \infty$, $\varphi(t,\omega,x) \le B$ for all $(t,\omega,x) \in [0,T] \times \bar{\mathcal{M}} \times \mathcal{Y}$.*

In later sections we will set $T = 1$, and hence the dependence of $\bar{\mathcal{A}}$ and $\bar{\mathcal{A}}_b$ on T can be omitted. Let $M_b(\mathcal{M})$ denote the space of bounded Borel measurable functions on $\mathcal{M}$. We then have the following representation formula for Poisson random measures.

Theorem 3.3.3. *Let $F \in M_b(\mathcal{M})$. Then for any $\theta > 0$,*

$$-\log \mathbb{E}_\theta [\exp(-F(N))] = \inf_{\varphi \in \bar{\mathcal{A}}_b} \bar{\mathbb{E}} \left[\theta L_T(\varphi) + F(N^{\theta\varphi}) \right]. \tag{3.3.3}$$

Proof. For $F \in M_b(\mathcal{M})$ and $\theta > 0$, it follows from Theorem 2.1 of [6] that

$$\begin{aligned} -\log \mathbb{E}_\theta [\exp(-F(N))] &= -\log \bar{\mathbb{E}} \left[\exp\left(-F(N^\theta)\right) \right] \\ &= \inf_{\varphi \in \bar{\mathcal{A}}} \bar{\mathbb{E}} \left[\theta L_T(\varphi) + F(N^{\theta\varphi}) \right]. \end{aligned}$$

Moreover, Theorem 2.4 of [5] states that the above infimization can in fact be taken over the smaller class of controls, such that $\varphi(t,\omega,x)$ is bounded by $B < \infty$ for all $(t,\omega) \in [0,T] \times \bar{\mathcal{M}}$ and for all x within some compact set, and $\varphi(t,\omega,x)$ is identical to 1 for x outside the set. Since $\bar{\mathcal{A}}_b \subset \bar{\mathcal{A}}$ contains this class of controls, we obtain (3.3.3). $\square$

3.3.2 Variational representation for the empirical measure process

In this section we derive a variational representation formula for the empirical measure process μ^n. We represent μ^n as a solution to a stochastic differential equation that is driven by finitely many i.i.d Poisson random measures, and using thinning functions to obtain the desired jump rates. We then derive a variational representation formula for μ^n, by viewing it as the image of a measurable mapping that acts on the collection of rescaled Poisson random measures.

Take $\nu = m$, so that $\nu_T = m_T \otimes m$. For $n \in \mathbb{N}$, let $\{N_v^n, v \in \mathcal{V}\}$ be a collection of i.i.d Poisson random measures (on $\mathcal{Y}_T$) with intensity measure $n\nu_T$. Thus we have the following SDE representation for the empirical measure process: for $t \in [0,T]$,

$$\mu^n(t) = \mu^n(0) + \sum_{v \in \mathcal{V}} v \int_{[0,t)} \int_{\mathcal{Y}} 1_{[0,\lambda_v^n(\mu^n(s))]}(x) \frac{1}{n} N_v^n(dsdx). \tag{3.3.4}$$

The existence of a solution to (3.3.4) is explained in the following argument.

From Condition 2.3.1 and (2.3.1), we can set

$$M' = \sup_{v \in \mathcal{V}, x \in \mathcal{S}_n, n \in \mathbb{N}} \lambda_v^n(x) < \infty. \tag{3.3.5}$$

Let $\mathcal{M}_{atom}$ denote the set of all $\mathrm{m} = \{\mathrm{m}_v, v \in \mathcal{V}\}$, where for each $v \in \mathcal{V}$, m_v is an atomic measure on $\mathcal{Y}_T$, with the property that $\mathrm{m}_v(\{t\} \times [0, M']) > 0$ for only finitely many t. Define $h : \mathcal{M}_{atom} \times \mathcal{S} \times \left([0,\infty)^{\mathcal{S}_n}\right)^{\otimes|\mathcal{V}|} \to D([0,1] : \mathcal{S})$ as the mapping that takes $(\mathrm{m}, \rho, \lambda^n) \in \mathcal{M}_{atom} \times \mathcal{S} \times \left([0,\infty)^{\mathcal{S}_n}\right)^{\otimes|\mathcal{V}|}$ to the process y defined by

$$y\left(t\right) = \rho + \sum_{v \in \mathcal{V}} v \int_{[0,t)} \int_{\mathcal{Y}} 1_{[0,\lambda_v^n(y(s))]}(x) \mathrm{m}_v(dsdx). \tag{3.3.6}$$

The existence of a solution $y(\cdot)$ to (3.3.6) is easily verified by the following recursive construction. Set $t_0 = 0$, and define $y^0\left(t\right) = \rho$ for $t \geq 0$. Assume as part of the recursive construction that for some $k \in \mathbb{N}_0$, a solution $y^k\left(\cdot\right)$ to (3.3.6) has been constructed on the interval $[0, t_k]$, and that $y^k\left(t\right) = y^k\left(t_k\right)$ for $t \geq t_k$. For any $t \in [t_k, T]$ and $v \in \mathcal{V}$, let

$$A_v\left(t\right) = \left\{(s,x) : s \in [t_k, t], x \in \left[0, \lambda_v^n\left(y^k(s)\right)\right]\right\},$$

and

$$t_{k+1} = \inf\left\{t > t_k \text{ such that for some } v \in \mathcal{V}, \mathrm{m}_v\left(A_v\left(t\right)\right) > 0\right\} \wedge T.$$

We then define $y^{k+1} : [0,T] \to \mathbb{R}^d$ by setting $y^{k+1}\left(t\right) = y^k\left(t\right)$ for $t \in [0, t_{k+1})$,

$$y^{k+1}\left(t_{k+1}\right) = y^k\left(t_k\right) + \sum_{v \in \mathcal{V}} v \int_{[t_k,t_{k+1}]} \int_{\mathcal{Y}} 1_{[0,\lambda_v^n\left(y^k(s)\right)]}(x) \mathrm{m}_v(dsdx),$$

and setting $y^{k+1}\left(t\right) = y^{k+1}\left(t_{k+1}\right)$ for $t \in [t_{k+1}, T]$.

Since m_v has finitely many atoms on $[0,T] \times [0, M']$, the construction will produce a function defined on all of $[0,T]$ in $L < \infty$ steps, at which time we set $y\left(t\right) = y^L\left(t\right)$. Since $N \in \mathcal{M}_{atom}$ for $\mathbb{P}_n-$a.e. ω, we can write

$$\mu^n\left(t,\omega\right) = h\left(\frac{1}{n}N^n, \mu^n(0,\omega), \lambda^n\right)(t) \tag{3.3.7}$$

for $\mathbb{P}_n-$a.e. $\omega \in \mathcal{M}$.

We now describe two classes of controls that will be used below. Recall that $\overline{\nu}_T = \nu_T \otimes m$. Let $\left\{\bar{N}_v^n, v \in \mathcal{V}\right\}$ be a collection of i.i.d Poisson random measures on $\mathcal{W}_T$ with intensity measure $n\overline{\nu}_T$. Let

$$\mathcal{G}_t^n = \sigma\left\{\bar{N}^n\left((0,s] \times A\right) : 0 \leq s \leq t, A \in \mathcal{B}\left(\mathcal{Y}\right)\right\}$$

and let $\mathcal{F}_t^n$ denote its completion under $\bar{\mathbb{P}}$. Denote by $\bar{\mathcal{P}}^n$ the predictable σ-field on $[0,T] \times \bar{\mathcal{M}}$ with the filtration $\{\mathcal{F}_t^n : 0 \leq t \leq T\}$ on $\left(\bar{\mathcal{M}}, \mathcal{B}\left(\bar{\mathcal{M}}\right)\right)$. For $n \in \mathbb{N}$, we denote $\bar{\mathcal{A}}_b^{\otimes n}$ and $\bar{\mathcal{A}}^{\otimes n}$ as the $n-$fold Cartesian product of $\bar{\mathcal{A}}_b$ and $\bar{\mathcal{A}}$. Given $\varphi \in \bar{\mathcal{A}}_b^{\otimes|\mathcal{V}|}$ we define the controlled jump Markov process $\hat{\mu}^n \in D\left([0,T] : \mathcal{S}\right)$ to be the solution to the following SDE: for $t \in [0,T]$,

$$\hat{\mu}^n\left(t\right) = \mu^n(0) + \sum_{v \in \mathcal{V}} v \int_{[0,t)} \int_{\mathcal{Y}} 1_{[0,\lambda_v^n(\hat{\mu}^n(s))]}(x) \int_{[0,\infty)} 1_{[0,\varphi_v(s,x)]}(r) \frac{1}{n} \bar{N}_v^n(dsdxdr). \tag{3.3.8}$$

As described previously, $\varphi_v(s,x)$ will control the jump rate as a function of (s,ω,x). In particular, the overall jump rate is the product $\lambda_v^n\left(\hat{\mu}^n(s)\right)\varphi_v(s,x)$, so that $\varphi_v(s,x)$ perturbs the jump rate away from that of the original model. For $v \in \mathcal{V}$, let $N_v^{n\varphi}$ be defined as in (3.3.1), with φ replaced by φ_v and $\bar{N}$ replaced by $\bar{N}_v^n$, and let $N^{n\varphi} = \{N_v^{n\varphi}, v \in \mathcal{V}\}$. For fixed $\varphi \in \bar{\mathcal{A}}_b^{\otimes|\mathcal{V}|}$, $N^{n\varphi} \in \mathcal{M}_{atom}$ a.s. From the definition of $h(\cdot)$, it is clear that (3.3.8) is equivalent to the relation

$$\hat{\mu}^n = h\left(\frac{1}{n}N^{n\varphi}, \mu^n(0), \lambda^n\right).$$

Applying Theorem 3.3.3 and (3.3.7) we obtain the following representation formula for μ^n.

Lemma 3.3.4. *For $F \in M_b(\mathcal{S})$,*

$$\begin{aligned} & -\frac{1}{n}\log \mathbb{E}\left[\exp(-nF(\mu^n))\right] \\ = & \inf_{\varphi \in \bar{\mathcal{A}}_b^{\otimes|\mathcal{V}|}} \bar{\mathbb{E}}\left[\sum_{v\in\mathcal{V}} L_T\left(\varphi_v\right) + F(\hat{\mu}^n) : \hat{\mu}^n = h\left(\frac{1}{n}N^{n\varphi}, \mu^n(0), \lambda^n\right)\right]. \end{aligned}$$

Proof. Consider the space $\tilde{\mathcal{Y}} = \mathcal{Y} \times \mathcal{V}$ and define a collection of Poisson random measures $\{\tilde{N}^n\}$ on $\tilde{\mathcal{Y}}$ with intensity measure $n\nu_T \otimes |\cdot|$, where $|\cdot|$ is the counting measure on $\mathcal{V}$. Then the SDE representation (3.3.4) is equivalent to

$$\mu^n(t) = \mu^n(0) + \int_{[0,t)}\int_{\tilde{\mathcal{Y}}} v 1_{[0,\lambda_v^n(\mu^n(s))]}(x)\frac{1}{n}\tilde{N}^n(dsdxdv), \;\; t\in[0,T].$$

Therefore for any $F \in M_b(\mathcal{S})$, applying Theorem 3.3.3 with $\mathcal{Y}$ replaced by $\tilde{\mathcal{Y}}$, ν replaced by $n\nu_T \otimes |\cdot|$, $F(\cdot)$ replaced by $nF \circ h\left(\frac{1}{n}\cdot, \mu^n(0), \lambda^n\right)$ yields

$$\begin{aligned} -\frac{1}{n}\log \mathbb{E}\left[\exp(-nF(\mu^n))\right] &= -\frac{1}{n}\log\mathbb{E}\left[\exp\left(-nF\circ h\left(\frac{1}{n}N^n, \mu^n(0), \lambda^n\right)\right)\right] \\ &= \inf_{\varphi \in \bar{\mathcal{A}}_b^{\otimes|\mathcal{V}|}} \bar{\mathbb{E}}\left[\sum_{v\in\mathcal{V}} L_T\left(\varphi_v\right) + F\circ h\left(\frac{1}{n}N^{n\varphi}, \mu^n(0), \lambda^n\right)\right]. \end{aligned}$$

□

We now derive a simpler form of the variational representation formula than the one given in Lemma 3.3.4. The starting point for Lemma 3.3.4 is the representation given in [6], which is general enough to cover situations where the different points in $\mathcal{Y}$ correspond to different "types" of jumps. For our purposes this is in fact more general than we need, since all points in $\mathcal{Y}$ correspond to exactly the same type of jump, and all that is needed from the space $\mathcal{Y}$ is that it be big enough that arbitrary jump rates [such as $\lambda_v^n\left(\hat{\mu}^n(s)\right)$] can be obtained by thinning. Because all the x's play an identical role, one expects, and we will verify using Jensen's inequality, that one can restrict to controls with no x-dependence. Thus we will replace the time, state and v dependent controls $\bar{\mathcal{A}}_b^{\otimes|\mathcal{V}|}$ by controls $\mathcal{A}_b^{\otimes|\mathcal{V}|}$ that only have time and v dependence, and rewrite the running cost as a function of the new controlled jump rates.

Definition 3.3.5. *Define $\mathcal{A}_b$ to be the class of $\bar{\mathcal{P}}\backslash\mathcal{B}[0,\infty)$ finite measurable maps $\varphi : [0,T] \times \bar{\mathcal{M}} \to [0,\infty)$.*

Define $\Lambda^n : \mathcal{A}_b^{\otimes|\mathcal{V}|} \times \mathcal{S} \to D\left([0,1] : \mathcal{S}\right)$ by

$$\Lambda^n\left(\bar{\alpha},\rho\right)(t) = \rho + \sum_{v\in\mathcal{V}} v \int_{[0,t)} \int_{\mathcal{Y}} 1_{[0,\bar{\alpha}_v(s)]}(x) \frac{1}{n} N_v^n(dsdx). \tag{3.3.9}$$

$\Lambda^n\left(\cdot,\rho\right)$ is well-defined for $\bar{\alpha} \in \mathcal{A}_b^{\otimes|\mathcal{V}|}$.

We are now in a position to state the main variational representation formula, the proof of which is deferred to Appendix B. The representation is the right one for finite state Markov chains, and expresses the variational integrand as the sum of a cost for perturbing jump rates, plus the expected value of the test function evaluated at the process which uses these perturbed rates.

Theorem 3.3.6. *Let $F \in M_b\left(\mathcal{S}\right)$. Then*

$$\begin{aligned} &-\frac{1}{n}\log \mathbb{E}\left[\exp(-nF(\mu^n))\right] \\ &= \inf_{\bar{\alpha}\in\mathcal{A}_b^{\otimes|\mathcal{V}|}} \bar{\mathbb{E}}\left[\sum_{v\in\mathcal{V}} \int_0^1 \lambda_v^n\left(\bar{\mu}^n(t)\right) \ell\left(\frac{\bar{\alpha}_v(t)}{\lambda_v^n\left(\bar{\mu}^n(t)\right)}\right) dt + F(\bar{\mu}^n) : \bar{\mu}^n = \Lambda^n\left(\bar{\alpha},\mu^n\left(0\right)\right)\right]. \end{aligned}$$

3.3.3 The Law of Large Numbers limit

We prove the law of large numbers limit (Theorem 2.3.2) at the end of this section. First recall the law of large numbers result for scaled Poisson random measures: for any $A \in \mathcal{B}([0,1] \times [0,\infty))$ such that $m_1 \otimes m(A) < \infty$, $\frac{1}{n} N_v^n(A) \to m_1 \otimes m(A)$ in probability, for any $v \in \mathcal{V}$. This implies that for any $f \in C_c([0,1] \times [0,\infty))$ we have $\int_0^1 \int_0^\infty f(s,x) \frac{1}{n} N_v^n(dsdx) \to \int_0^1 \int_0^\infty f(s,x)\, dsdx$. Rewrite (3.3.4) as

$$\mu^n(t) = \rho_0 + \sum_{v \in \mathcal{V}} v \int_{[0,t)} \lambda_v^n(\mu^n(s))\, ds + M^n(t),$$

where $M^n(t) = \sum_{v \in \mathcal{V}} \int_0^t \int_0^\infty 1_{[0,\lambda_v^n(\mu^n(s))]}(x)(\frac{1}{n} N_v^n(dsdx) - dsdx)$. Recall that λ_v^n converges uniformly to λ_v (by Condition 2.3.1). Therefore if for any $\varepsilon > 0$, $\mathbb{P}(\sup_{t \in [0,T]} \|M^n(t)\| > \varepsilon) \to 0$, then $\mu^n \to \mu$ in probability (uniformly on $t \in [0,T]$), for some μ which satisfies

$$\mu(t) = \rho_0 + \sum_{v \in \mathcal{V}} v \int_{[0,t)} \lambda_v(\mu(s))\, ds,$$

which is the integral version of (2.3.2). The uniqueness of the solution $\mu(\cdot)$ to the above equation follows from the Lipschitz continuity of $\lambda_v(\cdot)$.

Since $\{M^n(t), t \geq 0\}$ is an $\{\mathcal{F}_t^n\}-$martingale, for any $\varepsilon > 0$ Doob's inequality gives

$$
\begin{aligned}
& \mathbb{P}\left(\sup_{t\in[0,T]}\|M^n(t)\|>\varepsilon\right) \\
\leq\ & \frac{1}{\varepsilon^2}\mathbb{E}\left\|M^n(T)\right\|^2 \\
=\ & \frac{1}{\varepsilon^2}\mathbb{E}\left\|\sum_{v\in\mathcal{V}}\int_0^T\int_0^\infty 1_{[0,\lambda_v^n(\mu^n(s))]}(x)\frac{1}{n}(N_v^n(dsdx)-ndsdx)\right\|^2 \\
\leq\ & \frac{|\mathcal{V}|}{n\varepsilon^2}\mathbb{E}\int_0^1\int_0^M dsdx \\
\to\ & 0
\end{aligned}
$$

as $n\to\infty$, where M is defined as in (2.3.1). This completes the proof.

3.4 Proof of the LDP Upper Bound

A large deviation upper bound for a general class of Markov processes was obtained in [13]. We briefly describe the results in [13] below, specialized to the current setting. For every $n\in\mathbb{N}$, recall the Markov process μ^n on $\mathcal{S}_n$ with infinitesimal generator $\mathcal{L}_n$ given by (2.2.5). Theorem 1.1 of [13] applies to a slightly different class of Markov processes, with infinitesimal generator described by

$$
\mathcal{L}_n^0(f)(x)=n\sum_{v\in\mathcal{V}}\lambda_v(x)\left[f\left(x+\frac{1}{n}v\right)-f(x)\right], \tag{3.4.1}
$$

where λ_v takes values in (2.3.6).

The difference between (2.2.5) and (3.4.1) is the $n-$dependence of jump rates. However, since λ_v^n convergences uniformly to λ_v, Theorem 3.4.1 below implies that these processes have the same LDP rate function. The proof follows by a standard coupling argument, and will be deferred to Appendix C.

We assume Condition 2.3.1 for the results in this section. In what follows we denote $\|x-y\|_\infty = \sup_{t\in[0,1]} \|x(t) - y(t)\|$ for $x, y \in D([0,1] : \mathcal{S})$. Recall that two squences of Markov processes $\{X^n\}_{n\in\mathbb{N}}, \{Y^n\}_{n\in\mathbb{N}} \subset D([0,1] : \mathcal{S})$ are called *exponentially equivalent*, if for each $\varepsilon > 0$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\|X^n - Y^n\|_\infty > \varepsilon\right) = -\infty.$$

Exponential equivalence implies that $\{X^n\}_{n\in\mathbb{N}}$ satisfies the LDP with a given rate function if and only if $\{Y^n\}_{n\in\mathbb{N}}$ satisfies the LDP with the same rate function (Theorem 1.3.3 of [12]).

Theorem 3.4.1. *Assume the family of jump rates $\{\lambda_v(x), x \in \mathcal{S}, v \in \mathcal{V}\}$ satisfies Conditions 2.3.1. Let $\{X^n\}_{n\in\mathbb{N}}, \{Y^n\}_{n\in\mathbb{N}}$ be a squence of Markov processes with generator $\mathcal{L}_n$ and $\mathcal{L}_n^0$ respectively, and with $X^n(0) = Y^n(0)$. Then $\{X^n\}$ and $\{Y^n\}$ are exponentially equivalent.*

For $x, \alpha \in \mathbb{R}^d$, define

$$H(x,\alpha) = \sum_{v\in\mathcal{V}} \lambda_v(x) \left(\exp\langle\alpha, v\rangle - 1\right).$$

Note that H is continuous. Let L^0 be its Legendre-Fenchel transform defined by

$$L^0(x,\beta) = \sup_{\alpha\in\mathbb{R}^d} \left[\langle\alpha,\beta\rangle - H(x,\alpha)\right]. \tag{3.4.2}$$

Also, for $t \in [0,1]$ define I_t^0 as in (3.1.5), but with L replaced by L^0.

Proposition 3.4.2. *For any compact set $K \subset \mathcal{S}$ and $M < \infty$, the set*

$$\left\{\gamma : I^0(\gamma) \leq M, \gamma(0) \in K\right\}$$

is compact. Assume the family of jump rates $\{\lambda_v(x), x \in \mathcal{S}, v \in \mathcal{V}\}$ satisfies Conditions 2.3.1. Also, assume that the initial conditions $\{\mu^n(0)\}_{n\in\mathbb{N}}$ are deterministic, and $\mu^n(0) \to \mu_0 \in P(\mathcal{X})$ as n tends to infinity. Let $\{Y^n\}_{n\in\mathbb{N}}$ be a sequence of Markov processes with generator $\mathcal{L}_n^0$, and $Y^n(0) = \mu^n(0)$. Then $\{Y^n\}$ satisfies the large deviation upper bound with rate function I^0.

Proof. This follows from Theorem 1.1 of [13] with $\varepsilon = 1/n$, $a(\cdot) = b(\cdot) = 0$, and $\mu_x(\cdot) = 1_{x\in\mathcal{S}} \sum_{v\in\mathcal{V}} \lambda_v(x)\, \delta_v(\cdot)$. □

We have introduced functions L^0 [in (3.4.2)] and L [in (3.1.3)], defined respectively in terms of a Legendre transform and relative entropy. The next proposition shows they are two representations of the same local rate function.

Proposition 3.4.3. *$L(x,\beta) = L^0(x,\beta)$ for all $x \in \mathcal{S}, \beta \in \Delta^{d-1}$.*

Proof. Defining $h_{v,a} : \mathbb{R}^d \to \mathbb{R}$ by $h_{v,a}(\alpha) = a(\exp(\langle\alpha, v\rangle) - 1)$ for $v \in \mathbb{R}^d$ and $a \in [0,\infty)$, we can write $H(x,\alpha) = \sum_{v\in\mathcal{V}} h_{v,\lambda_v(x)}(\alpha)$. The Legendre-Fenchel transform of $h_{v,a}$ can be computed explicitly as

$$h_{v,a}^*(\beta) = \begin{cases} a\ell(y) & \text{if } \beta = avy, \\ \infty & \text{otherwise.} \end{cases}$$

Since H is a finite sum of convex functions, we can apply a standard result in convex analysis to calculate its Legendre-Fenchel transform (see, e.g., Theorem D.4.2 of

[12]):

$$\left(\sum_{v\in\mathcal{V}} h_{v,\lambda_v(x)}\right)^*(\beta) = \inf\left\{\sum_{v\in\mathcal{V}} h^*_{v,\lambda_v(x)}(\beta_v) : \sum_{v\in\mathcal{V}} \beta_v = \beta\right\}.$$

Hence,

$$L^0(x,\beta) = \inf_{q:\sum_{v\in\mathcal{V}} vq_v=\beta} \sum_{v\in\mathcal{V}} \lambda_v(x)\,\ell\left(\frac{q_v}{\lambda_v(x)}\right) = L(x,\beta).$$

□

Combining Theorem 3.4.1, Proposition 3.4.2 and Proposition 3.4.3, we conclude that that $I(\gamma)$ satisfies (3.1.7), and has compact level sets for compact sets of initial conditions.

3.5 Properties of the Local Rate Function

In this section we establish useful properties of the function

$$L(x,\beta) = \inf_{q:\sum_{v\in\mathcal{V}} vq_v=\beta} \sum_{v\in\mathcal{V}} \lambda_v(x)\,\ell\left(\frac{q_v}{\lambda_v(x)}\right), \quad x\in\mathcal{S}, \beta\in\Delta^{d-1}, \tag{3.5.1}$$

which is defined in (3.1.3) as a proposed local rate function for $\{\mu^n\}_{n\in\mathbb{N}}$. In what follows we denote the relative interior of $\mathcal{S}$ by int($\mathcal{S}$). Also, for $a>0$, define

$$\begin{aligned} \mathcal{S}^a &\doteq \{p\in\mathcal{S} : \mathrm{dist}(p,\partial\mathcal{S}) \geq a\} \\ &= \left\{p\in\mathcal{S} : \inf_{x\in\partial\mathcal{S}} \|p-x\| \geq a\right\}. \end{aligned} \tag{3.5.2}$$

Note that for $x\in\mathcal{S}^a$, $x_i \geq a/\sqrt{d}$ for $i=1,...,d$.

Throughout this section, we assume Conditions 2.3.1, 3.1.1 and 3.1.4 hold for

$\{\lambda_v(\cdot)\}_{v\in\mathcal{V}}$. Given a set of vectors $\{u_j\}_{j=1}^F \subset \mathbb{R}^d$, the positive cone spanned by $\{u_j\}_{j=1}^F$ is defined by

$$\mathcal{C}\{u_j\} \doteq \left\{ v \in \mathbb{R}^d : \text{there exist } a_j \geq 0 \text{ with } v = \sum_{j=1}^F a_j u_j \right\}. \tag{3.5.3}$$

Also, define $\mathcal{V}_x = \{v \in \mathcal{V} : \lambda_v(x) > 0\}$,
$\mathcal{V}_+ = \{v \in \mathcal{V} : \text{for any } a > 0, \ \inf_{x\in\mathcal{S}^a} \lambda_v(x) > 0\}$. Under Condition 3.1.1, for $x \in \operatorname{int}(\mathcal{S})$, $\mathcal{V}_x = \mathcal{V}_+$. The following observation follows from the definition of a communicating path.

Proposition 3.5.1. *For $x \in \operatorname{int}(\mathcal{S})$, $\mathcal{C}\{v : v \in \mathcal{V}_x\} = \mathcal{C}\{v : v \in \mathcal{V}_+\} = \Delta^{d-1}$.*

It is also convenient to introduce another function $\bar{L} : [0,\infty)^{|\mathcal{V}|} \times \Delta^{d-1} \to \mathbb{R}$ which, for any vectors $u \in [0,\infty)^{|\mathcal{V}|}$ and $\beta \in \Delta^{d-1}$, is defined by

$$\bar{L}(u,\beta) = \inf_{q:\sum_{v\in\mathcal{V}} vq_v = \beta} \sum_{v\in\mathcal{V}} u_v \ell\left(\frac{q_v}{u_v}\right). \tag{3.5.4}$$

Lemma 3.5.2. *$\bar{L}$ is jointly strictly convex in $u \in [0,\infty)^{|\mathcal{V}|}$ and $\beta \in \Delta^{d-1}$. The function L defined in (3.5.1) is nonnegative and uniformly continuous on compact subsets of $\operatorname{int}(\mathcal{S}) \times \Delta^{d-1}$, and for each $x \in \mathcal{S}$, $L(x,\cdot)$ is a strictly convex on Δ^{d-1}.*

Proof. The joint convexity of the function
$(u,q) \in [0,\infty)^{|\mathcal{V}|} \times [0,\infty)^{|\mathcal{V}|} \mapsto \sum_{v\in\mathcal{V}} u_v \ell(q_v/u_v)$ is immediate. The joint convexity of $\bar{L}$ then follows since it is the infimum of a jointly convex function subject to an affine constraint. To show that $\bar{L}(x,\cdot)$ is strictly convex, note that $q \mapsto \sum_{v\in\mathcal{V}} u_v \ell(q_v/u_v)$ is strictly convex, and goes to infinity as $\|q\| \to \infty$. Thus the minimum on a closed convex set is uniquely attained: given any $u \in [0,\infty)^{|\mathcal{V}|}$ and $\beta \in \Delta^{d-1}$, there exists $q^*(\beta)$ such that $\sum_{v\in\mathcal{V}} vq_v^* = \beta$, and $\bar{L}(u,\beta) = \sum_{v\in\mathcal{V}} u_v \ell(q_v^*/u_v)$. Therefore for any

$\beta_1 \neq \beta_2$ and $\delta \in (0,1)$,

$$\begin{aligned}
\delta \bar{L}(u,\beta_1) + (1-\delta)\bar{L}(u,\beta_2) &= \delta \sum_{v\in\mathcal{V}} u_v \ell\left(\frac{q_v^*(\beta_1)}{u_v}\right) + (1-\delta)\sum_{v\in\mathcal{V}} u_v \ell\left(\frac{q_v^*(\beta_2)}{u_v}\right) \\
&> \sum_{v\in\mathcal{V}} u_v \ell\left(\frac{\delta q_v^*(\beta_1) + (1-\delta) q_v^*(\beta_2)}{u_v}\right) \\
&\geq \bar{L}(u, \delta\beta_1 + (1-\delta)\beta_2).
\end{aligned}$$

The nonnegativity of L follows directly from the definition (3.5.1). For any $a > 0$, by Remark 3.1.2, $\lambda_v(\cdot)$ is either identically zero or uniformly bounded below away from zero on S^a. Fix any $\beta \in \Delta^{d-1}$. By Proposition 3.5.1, there exists $q \in [0,\infty)^{|\mathcal{V}|}$ such that $\sum_{v\in\mathcal{V}_+} v q_v = \beta$, and $q_v = 0$ for $v \in \mathcal{V}\backslash\mathcal{V}_+$. Since $\lambda_v(\cdot)$ is continuous, it follows that L is uniformly continuous on compact subsets of $\mathrm{int}(\mathcal{S}) \times \Delta^{d-1}$. The strict convexity claim follows by noting $L(x,\cdot) = \bar{L}(\lambda(x),\cdot)$ for any $x \in \mathcal{S}$. $\square$

The following elementary inequality can be proved using Legendre transforms.

Lemma 3.5.3. *For $a, q \in [0,\infty)$ we have $a\ell\left(\frac{q}{a}\right) + a(e-1) \geq q$.*

We now study the asymptotic behavior of the proposed local rate function L.

Proposition 3.5.4. *Given $a > 0$, there exist constants $B = B(a) < \infty$ and $C_2 = C_2(B), C_3 = C_3(B) < \infty$, such that*

$$L(x,\beta) \leq \begin{cases} C_2 \|\beta\| \log \|\beta\| & \text{if } x \in \mathcal{S}^a \text{ and } \beta \in \Delta^{d-1}, \|\beta\| > B \\ C_3 & \text{if } x \in \mathcal{S}^a \text{ and } \beta \in \Delta^{d-1}, \|\beta\| \leq B. \end{cases}$$

For $B < \infty$ sufficiently large, there exists $c_1 = c_1(B) > 0$ such that

$$L(x,\beta) \geq c_1 \|\beta\| \log \|\beta\|$$

if $x \in \mathcal{S}$ and $\beta \in \Delta^{d-1}$, $\|\beta\| > B$. In particular, $L(x,\beta)$ is superlinear in β, uniformly in x.

Proof. Fix $a > 0$. For any $B < \infty$, since $\left\{(x,\beta) \in \mathcal{S}^a \times \Delta^{d-1} : \|\beta\| \le B\right\}$ is a compact subset of $\mathrm{int}(\mathcal{S}) \times \Delta^{d-1}$, the uniform boundedness of L on this set follows directly from Lemma 3.5.2.

For the upper bound when $\|\beta\| > B$, we first assume $\|\beta\| = 1$. By Proposition 3.5.1, there exists a bounded vector $q = q(\beta) \in [0,\infty)^{|\mathcal{V}|}$, such that $\sum_{v \in \mathcal{V}_+} v q_v = \beta$, and $q_v = 0$ for $v \in \mathcal{V}\backslash\mathcal{V}_+$. By taking an open cover on $\left\{\beta : \beta \in \Delta^{d-1}, \|\beta\| = 1\right\}$, one can assume $\max_{v, \|\beta\|=1} |q_v(\beta)|$ is bounded. By scaling, it follows that there exists some constant $c_0 > 0$, such that for any $\beta \in \Delta^{d-1}$, there exists a vector $q \in [0,\infty)^{|\mathcal{V}|}$ such that $\sum_{v \in \mathcal{V}_+} v q_v = \beta$, $\max_v |q_v| \le c_0 \|\beta\|$, and $q_v = 0$ for $v \in \mathcal{V}\backslash\mathcal{V}_+$. It follows that for some $c_4 < \infty$,

$$L(x,\beta) \le c_4 \sum_{v \in \mathcal{V}} q_v \log \frac{q_v}{\lambda_v(x)} \le C_2 \|\beta\| \log \|\beta\|$$

if $\|\beta\| \ge B$, for some B sufficiently large and $x \in \mathcal{S}^a$. This finishes the proof of the upper bound.

Now, consider the lower bound in $\{(x,\beta) : \|\beta\| > B, x \in \mathcal{S}\}$. Using Proposition 3.4.3, we have for $t > 0$, $a = t\frac{\beta}{\|\beta\|}$, and $M < \infty$ defined as in (2.3.1),

$$\begin{aligned} L(x,\beta) &\ge \langle a,\beta\rangle - H(x,a) \\ &\ge t\|\beta\| - \sum_{v \in \mathcal{V}} \lambda_v(x) \exp\langle a, v\rangle \\ &\ge t\|\beta\| - M|\mathcal{V}| \exp\left(\max_{v \in \mathcal{V}} \|v\|\, t\right). \end{aligned}$$

Setting $t = \frac{1}{\max_{v\in\mathcal{V}}\|v\|} \log\|\beta\|$, this implies

$$L(x,\beta) \geq \frac{1}{\max_{v\in\mathcal{V}}\|v\|} \|\beta\| \log\|\beta\| - M|\mathcal{V}| \|\beta\| \geq c_1 \|\beta\| \log\|\beta\|,$$

for some constant $c_1 > 0$, provided $\|\beta\|$ is sufficiently large. $\square$

When applying Proposition 3.5.4, we will always assume (without mentioning in explicitly) that B is chosen to be a sufficiently large number so that all parts of Proposition 3.5.4 hold.

Proposition 3.5.5. *Given* $0 \leq a < b \leq 1$ *and* $\xi > 0$, *suppose that* $\gamma \in AC\left([a,b] : \mathcal{S}^{\xi}\right)$ *satisfies* $\int_a^b L(\gamma(s), \dot{\gamma}(s))\, ds < \infty$. *Let* $\left\{\gamma^{\delta}\right\}_{\delta>0} \in D\left([a,b] : \mathcal{S}^{\xi}\right)$ *be such that* $\sup_{t\in[a,b]} \left\|\gamma^{\delta}(t) - \gamma(t)\right\| \to 0$ *as* $\delta \to 0$. *Then for any* $\varepsilon > 0$, *there exists* $\delta_0 = \delta_0(\xi,\varepsilon) > 0$, *such that for* $\delta < \delta_0$, $\left|\int_a^b L(\gamma(s), \dot{\gamma}(s))\, ds - \int_a^b L\left(\gamma^{\delta}(s), \dot{\gamma}(s)\right) ds\right| < \varepsilon$.

Proof. Fix $\xi > 0$, $0 \leq a < b \leq 1$. Let $A_0 \subset [a,b]$ be the set for which $\dot{\gamma}(\cdot)$ is well defined, and let $B = B(\xi/2)$ be the sufficiently large constant from Proposition 3.5.4. Then $[a,b]\backslash A_0$ has measure 0. Define $A = \{s \in A_0 : \|\dot{\gamma}(s)\| \leq B\}$. Given $\varepsilon > 0$, choose B larger if necessary such that

$$\int_{[a,b]\backslash A} L(\gamma(s), \dot{\gamma}(s))\, ds \leq \varepsilon. \tag{3.5.5}$$

Then by dominated convergence and the continuity of $L(\cdot,\beta)$ for fixed $\beta \in \Delta^{d-1}$ established in Lemma 3.5.2, we have

$$\int_A L\left(\gamma^{\delta}(s), \dot{\gamma}(s)\right) ds \to \int_A L(\gamma(s), \dot{\gamma}(s))\, ds. \tag{3.5.6}$$

Then, with constants $C_2 = C_2(B)$, $c_1 = c_1(B)$ from Proposition 3.5.4, it follows that

$$\begin{aligned}
\int_{[a,b]\backslash A} L\left(\gamma^{\delta}(s), \dot{\gamma}(s)\right) ds &\leq C_2 \int_{[a,b]\backslash A} \|\dot{\gamma}(s)\| \log \|\dot{\gamma}(s)\| \, ds \\
&\leq C_2/c_1 \int_{[a,b]\backslash A} L(\gamma(s), \dot{\gamma}(s)) \, ds \\
&\leq C_2 \varepsilon / c_1.
\end{aligned}$$

Together with (3.5.5) and (3.5.6), this implies that for δ sufficiently small,

$$\left| \int_a^b L(\gamma(s), \dot{\gamma}(s)) \, ds - \int_a^b L\left(\gamma^{\delta}(s), \dot{\gamma}(s)\right) ds \right| \leq \varepsilon + \varepsilon + C_2\varepsilon/c_1 = (2 + C_2/c_1)\varepsilon.$$

□

Lemma 3.5.6. *Suppose that* $\gamma \in AC([0,1] : \mathcal{S})$ *satisfies* $\int_a^b L(\gamma(s), \dot{\gamma}(s)) \, ds < \infty$ *for some* $0 \leq a < b \leq 1$. *Then*

$$\|\gamma(t) - \gamma(a)\| \log \frac{1}{t-a} \to 0 \quad \text{as } t \downarrow a.$$

Proof. Fix $t \in (a,b)$ and let $A_0 \subset [a,b]$ be the set for which $\dot{\gamma}(\cdot)$ is well defined. Then $[a,b] \backslash A_0$ has measure 0. Now observe that $\|\dot{\gamma}(\cdot)\| \log \|\dot{\gamma}(\cdot)\| \in L^1([a,b] : \mathcal{S})$. Take some sufficiently large $B < \infty$ from Proposition 3.5.4, and let $A = \{s \in [a,b] : \|\dot{\gamma}(s)\| \leq B\}$. Then by Proposition 3.5.4, there exists $c_1 = c_1(B)$, such that

$$\begin{aligned}
\int_a^b \|\dot{\gamma}(s)\| \log \|\dot{\gamma}(s)\| \, ds &\leq \frac{1}{c_1} \int_{[a,b]\backslash A} L(\gamma(s), \dot{\gamma}(s)) \, ds + \int_A B \log B ds \\
&\leq \frac{1}{c_1} \int_a^b L(\gamma(s), \dot{\gamma}(s)) \, ds + (B \log B)(b-a).
\end{aligned}$$

By Jensen's inequality,

$$\int_a^t \|\dot{\gamma}(s)\| \log \|\dot{\gamma}(s)\| \, ds \geq (t-a) \left\| \frac{\gamma(t)-\gamma(a)}{t-a} \right\| \log \left\| \frac{\gamma(t)-\gamma(a)}{t-a} \right\|$$
$$= \|\gamma(t)-\gamma(a)\| \log \frac{\|\gamma(t)-\gamma(a)\|}{t-a}$$

The lemma follows by observing that both the left hand side and $\|\gamma(t)-\gamma(a)\| \log \|\gamma(t)-\gamma(a)\|$ goes to zero as $t \downarrow a$. $\square$

Lemma 3.5.7. *Let $c_0(\delta)$ be given such that $c_0(\delta) \to 1$ as $\delta \to 0$. Suppose that $x \in \mathcal{S}$, $\{x^\delta\}_{\delta>0} \in$int$(\mathcal{S})$, are such that $\|x - x^\delta\| \to 0$ as $\delta \to 0$, and for any $\delta > 0$ and $v \in \mathcal{V}$, $\frac{\lambda_v(x)}{\lambda_v(x^\delta)} \leq c_0(\delta)$. Then there exists $c = c(\delta)$ that only depends on $c_0(\delta)$ and $\|x - x^\delta\|$, and goes to zero as $\delta \to 0$, such that for any $\beta \in \mathcal{C}\{v : v \in \mathcal{V}_x\}$,*

$$L(x,\beta) \leq (1 + c(\delta)) L\left(x^\delta, \beta\right) + c(\delta).$$

Proof. Fix $\beta \in \mathcal{C}\{v : v \in \mathcal{V}_x\}$, and take $q \in [0,\infty)^{|\mathcal{V}|}$ such that $\sum_{v \in \mathcal{V}_x} v q_v = \beta$ and $q_v = 0$ for $v \in \mathcal{V} \backslash \mathcal{V}_x$. By the continuity of $\{\lambda_v(\cdot)\}$ and the fact that they are either identically zero or bounded uniformly from below in int$(\mathcal{S})$ (Condition 3.1.1), it follows that for $x \in$int$(\mathcal{S})$ and δ sufficiently small, $\mathcal{V}_{x^\delta} = \mathcal{V}_x$. For $x \in \partial\mathcal{S}$, $\mathcal{V}_x \subset \mathcal{V}_{x^\delta}$ if $\{x^\delta\}_{\delta>0}$ approximate x from the interior. Therefore for δ sufficiently small, the same vector q we take before satisfies $\sum_{v \in \mathcal{V}_{x^\delta}} v q_v = \beta$ and $q_v = 0$ for $v \in \mathcal{V} \backslash \mathcal{V}_{x^\delta}$. It suffices to show that for any $\delta > 0$,

$$\sum_{v \in \mathcal{V}_x} \lambda_v(x) \, \ell\left(\frac{q_v}{\lambda_v(x)}\right) \leq (1 + c(\delta)) \sum_{v \in \mathcal{V}_x} \lambda_v\left(x^\delta\right) \ell\left(\frac{q_v}{\lambda_v(x^\delta)}\right) + c(\delta), \qquad (3.5.7)$$

where $c(\delta)$ is independent of q.

Note that there exist $C_1 < \infty$, such that

$$\begin{aligned}
&\sum_{v\in\mathcal{V}_x} \lambda_v(x)\, \ell\left(\frac{q_v}{\lambda_v(x)}\right) - \sum_{v\in\mathcal{V}_x} \lambda_v\left(x^\delta\right) \ell\left(\frac{q_v}{\lambda_v\left(x^\delta\right)}\right) \\
&\quad = -\sum_{v\in\mathcal{V}_x} q_v \log\frac{\lambda_v\left(x^\delta\right)}{\lambda_v(x)} + \sum_{v\in\mathcal{V}_x}\left(\lambda_v(x) - \lambda_v\left(x^\delta\right)\right) \\
&\quad \le \log c_0(\delta) \sum_{v\in\mathcal{V}_x} q_v + C_1 \left\|x - x^\delta\right\| \\
&\quad \le \log c_0(\delta) \sum_{v\in\mathcal{V}_x} \left(\lambda_v\left(x^\delta\right) \ell\left(\frac{q_v}{\lambda_v\left(x^\delta\right)}\right) + \lambda_v\left(x^\delta\right)(e-1)\right) + C_1\left\|x - x^\delta\right\| \\
&\quad \le \log c_0(\delta) \sum_{v\in\mathcal{V}_x} \lambda_v\left(x^\delta\right) \ell\left(\frac{q_v}{\lambda_v\left(x^\delta\right)}\right) + C_2(\delta),
\end{aligned}$$

for $C_2(\delta) = \log c_0(\delta)\, M\, |\mathcal{V}|\,(e-1) + C_1\left\|x - x^\delta\right\|$, where the first inequality follows from the assumption and Lipschitz continuity of $\{\lambda_v(\cdot)\}$, and the second inequality follows from Lemma 3.5.3 by taking $a = \lambda_v\left(x^\delta\right)$ and $q = q_v$. It suffices to take $c(\delta) = \max\{\log c_0(\delta), C_2(\delta)\}$. $\square$

Notice that for $\beta \notin \mathcal{C}\{v : v \in \mathcal{V}_x\}$, the conclusion of the above Lemma holds trivially since the right hand side is infinity.

For $t \in [0,1]$ and $c > 0$, define

$$\gamma_c(s) = \gamma(cs), \quad s \in [0,t],$$

which is a time reparametrization of γ. The next result is used in the proof of the locally uniform LDP in Section 3.7. It states that given a path γ with finite cost, the cost of the path depends continuously on the reparametrization of time.

Proposition 3.5.8. *For $t \in [0,1)$ given, suppose $\gamma \in AC([0,1] : \mathcal{S})$ is such that $I_t(\gamma) < \infty$. Then the function $c \mapsto I_{t/c}(\gamma_c)$ is continuous at* 1.

Proof. First note that for c close to 1, $\gamma_c \in AC([0,t/c]:\mathcal{S})$ and

$$I_{t/c}(\gamma_c) = \int_0^{t/c} L(\gamma(cs), c\dot{\gamma}(cs))\, ds = \frac{1}{c}\int_0^t L(\gamma(r), c\dot{\gamma}(r))\, dr.$$

It suffices to bound the integral of $\frac{1}{c}L(\gamma, c\dot{\gamma}) - L(\gamma,\dot{\gamma})$. Recall the definition of L in (3.1.3). Since γ is absolutely continuous, $\dot{\gamma}(u)$ is a.s. well defined. For any fixed $u \in [0,t]$ such that $\dot{\gamma}(u)$ is well defined, for any $\varepsilon > 0$, there exists $q \in [0,\infty)^{|\mathcal{V}|}$, such that $\sum_{v\in\mathcal{V}} vq_v = c\dot{\gamma}(u)$ and

$$\sum_{v\in\mathcal{V}} \lambda_v(\gamma(u))\, \ell\left(\frac{q_v/c}{\lambda_v(\gamma(u))}\right) \le L(\gamma(u), \dot{\gamma}(u)) + \varepsilon.$$

We also have

$$\begin{aligned}
&\sum_{v\in\mathcal{V}} \lambda_v(\gamma(u))\, \ell\left(\frac{q_v/c}{\lambda_v(\gamma(u))}\right) \\
&\quad = \sum_{v\in\mathcal{V}} \left(\frac{q_v}{c}\log\frac{q_v/c}{\lambda_v(\gamma(u))} - \frac{q_v}{c} + \lambda_v(\gamma(u))\right) \\
&\quad = \frac{1}{c}\sum_{v\in\mathcal{V}} \left(q_v \log\frac{q_v}{\lambda_v(\gamma(u))} - q_v + \lambda_v(\gamma(u))\right) + \left(\frac{1}{c}\log\frac{1}{c}\right)\sum_{v\in\mathcal{V}} q_v \\
&\qquad + \left(1-\frac{1}{c}\right)\sum_{v\in\mathcal{V}} \lambda_v(\gamma(u)) \\
&\quad \ge \frac{1}{c}L(\gamma(u), c\dot{\gamma}(u)) + \left(\frac{1}{c}\log\frac{1}{c}\right)\sum_{v\in\mathcal{V}} q_v + \left(1-\frac{1}{c}\right)\sum_{v\in\mathcal{V}} \lambda_v(\gamma(u)).
\end{aligned}$$

Thus

$$\frac{1}{c}L(\gamma(u), c\dot{\gamma}(u)) - L(\gamma(u), \dot{\gamma}(u)) \le \varepsilon - \left(\frac{1}{c}\log\frac{1}{c}\right)\sum_{v\in\mathcal{V}} q_v - \left(1-\frac{1}{c}\right)\sum_{v\in\mathcal{V}} \lambda_v(\gamma(u)). \tag{3.5.8}$$

Similarly, by taking $q \in [0,\infty)^{|\mathcal{V}|}$, such that $\sum_{v\in\mathcal{V}} vq_v = \dot{\gamma}(u)$ and

$$\sum_{v\in\mathcal{V}} \lambda_v(\gamma(u))\,\ell\left(\frac{cq_v}{\lambda_v(\gamma(u))}\right) \le L(\gamma(u), c\dot{\gamma}(u)) + c\varepsilon,$$

an analogous computation yields

$$\frac{1}{c}L(\gamma(u), c\dot{\gamma}(u)) - L(\gamma(u), \dot{\gamma}(u)) \ge (\log c)\sum_{v\in\mathcal{V}} q_v - \left(1-\frac{1}{c}\right)\sum_{v\in\mathcal{V}} \lambda_v(\gamma(u)) - \varepsilon. \tag{3.5.9}$$

We apply Lemma 3.5.3 with $a = \lambda_v(\gamma(u))$ and $q = q_v/c$, and the boundedness of λ_v to obtain

$$\begin{aligned}\frac{1}{c}\sum_{v\in\mathcal{V}} q_v &\le \sum_{v\in\mathcal{V}}\left(\lambda_v(\gamma(u))\,\ell\left(\frac{q_v/c}{\lambda_v(\gamma(u))}\right) + \lambda_v(\gamma(u))(e-1)\right)\\ &\le L(\gamma(u), \dot{\gamma}(u)) + \varepsilon + c_1\end{aligned}$$

for some $c_1 < \infty$. Thus, combining with (3.5.8) and (3.5.9), we see that for c sufficiently close to 1,

$$\begin{aligned}&\left|\frac{1}{c}L(\gamma(u), c\dot{\gamma}(u)) - L(\gamma(u), \dot{\gamma}(u))\right|\\ &\quad\le M|\mathcal{V}|\left|1-\frac{1}{c}\right| + \max\left\{\log\frac{1}{c}, c\log c\right\}(L(\gamma(u), \dot{\gamma}(u)) + \varepsilon + c_1) + \varepsilon.\end{aligned}$$

Since $I(\gamma)$ is finite, one can integrate over $[0,t]$, take $c \to 1$, and then send $\varepsilon \to 0$ to complete the proof. $\square$

3.6 Proof of the LDP lower bound

We now turn to the proof of the LDP lower bound. It suffices to show that for any fixed trajectory $\gamma \in D([0,1] : \mathcal{S})$, given any $\varepsilon > 0$ and $\delta > 0$ there exists $\eta > 0$ such that if $\|\mu^n(0) - \gamma(0)\| < \eta$ for all n large enough,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\|\mu^n - \gamma\|_\infty < \delta\right) \geq -I(\gamma) - \varepsilon.$$

Without loss of generality we assume $I(\gamma) < \infty$.

One source of difficulty here is that the transition rates of μ^n may tend to zero as μ^n approaches the boundary of $\mathcal{S}$, which will lead to singularity of the local rate function. Our approach here adapts an idea from the study of a discrete time model in [15]. We first show that the singularity can be avoided except for $t = 0$, by slightly perturbing the original path, with arbitrarily small additional cost.

3.6.1 Perturbation argument

We start with a direct evaluation of the hitting probability of jump Markov processes on a finite state space.

Lemma 3.6.1. *Let $\{Y(t)\}_{t\geq 0}$ be a jump Markov process with finite state space $\{s_0, s_1, ..., s_n\}$. For $i = 0, ..., n-1$, suppose that the jump rate from state s_i to s_{i+1} is b_{i+1}, and the sum of jump rates from state s_i to all other states is bounded above by $c < \infty$. If $Y(0) = s_0$, then*

$$\mathbb{P}(Y(t) = s_n) \geq \frac{1}{n!}\left(\Pi_{i=1}^n b_i\right) t^n \exp(-ct).$$

Proof. Let $p(t)$ be the probability distribution of the process at time t: $p_i(t) = \mathbb{P}(Y(t) = s_i)$. Then the Kolmogorov forward equation takes the form $\dot{p} = Ap$, where A is the rate matrix for Y. By the comparison principle for ODEs, it follows that $p_i(t) \geq r_i(t)$ for all $i \in \mathcal{X}$, where $r(t)$ is the solution to

$$\begin{cases} \dot{r}_0 = -cr_0, \\ \dot{r}_i = b_i r_{i-1} - cr_i, \quad i = 1, ..., n, \\ r(0) = e_{s_0}. \end{cases}$$

Solving this equation explicitly gives $p_n(t) \geq r_n(t) = \frac{1}{n!}\left(\Pi_{i=1}^n b_i\right) t^n \exp(-ct)$. □

The idea of the perturbation argument is as follows. Recall the definition of $\mathcal{S}^a$ in (3.5.2). For any $a > 0$ fixed, by Remark 3.1.2, the rates $\lambda_v(\cdot)$ are either identically zero or uniformly bounded below away from zero within $\mathcal{S}^a$. Therefore, a standard approximation argument can be used to establish the LDP in $\mathcal{S}^a$, uniformly with respect to the initial condition. When $\gamma(0) = x \in \mathcal{S}/\mathcal{S}^a$, by using Condition 3.1.3, one can construct a perturbed trajectory of γ, that hits $\mathcal{S}^a$ in an arbitrarily short time as $a \to 0$, and in such a way that the difference in cost between γ and the perturbed trajectory can be made sufficiently small.

Lemma 3.6.2. *Assume the family of jump rates* $\{\lambda_v(x), x \in \mathcal{S}, v \in \mathcal{V}\}$ *satisfies Conditions 2.3.1, 3.1.1 and 3.1.3. Consider* $\gamma \in AC([0,1] : \mathcal{S})$ *such that* $I(\gamma) < \infty$. *Then given any* $\varepsilon > 0$, *there exists* $\tilde{b} > 0$, $D < \infty$ *and a trajectory* $\upsilon \in AC([0,1] : \mathcal{S})$ *such that*

i) $\upsilon(0) = \gamma(0)$ *and* $\|\upsilon - \gamma\|_\infty < \varepsilon$,

ii) $\upsilon_j(t) \geq \tilde{b} t^D$ *for* $j = 1, ..., d$ *and any* $t \in [0,1]$,

iii) $I(\upsilon) \leq I(\gamma) + \varepsilon$.

Proof. For $0 < \rho < 1$ define $\upsilon^{\rho} = \rho\mu + (1-\rho)\gamma$, where μ is the law of large numbers trajectory defined in (2.3.2) with $\mu(0) = \gamma(0)$. Let C_d be the diameter of $\mathcal{S}$. Then we have $\|\upsilon^{\rho} - \gamma\|_{\infty} = \rho\|\mu - \gamma\|_{\infty} \leq C_d\rho$. Also, one has the lower bound $\upsilon_j^{\rho}(t) \geq \rho\mu_j(t) \geq \rho b t^D$ for some $D < \infty$ by Condition 3.1.3. Therefore, it suffices to show iii). We first show that there exists $c(\rho) < \infty$ which goes to zero as $\rho \to 0$, such that for any $t \in [0,1]$,

$$L(\upsilon^{\rho}(t), \dot{\gamma}(t)) \leq (1 + c(\rho)) L(\gamma(t), \dot{\gamma}(t)) + c(\rho). \tag{3.6.1}$$

Indeed, since $\gamma_i(t)/\upsilon_i^{\rho}(t) \leq \frac{1}{1-\rho}$ and $\|\upsilon_i^{\rho}(t) - \gamma_i(t)\| \leq C_d\rho$ for $i = 1, ..., d$, by Condition 3.1.1

$$\frac{\lambda_v(\gamma(t))}{\lambda_v(\upsilon^{\rho}(t))} \leq \left(\frac{1}{1-\rho}\right)^{dC_1} \exp(f(\rho)), \text{ for every } v \in \mathcal{V},$$

where $f(\rho) = \max_{s\in[0,C_d\rho]} f_0(s)$. Since the right hand side goes to 1 as $\rho \to 0$, (3.6.1) follows by applying Lemma 3.5.7 with $x = \gamma(t)$ and $x^{\rho} = \upsilon^{\rho}(t)$.

Likewise, $\mu_i(t)/\upsilon_i^{\rho}(t) \leq 1/\rho$ and $\|\upsilon_i^{\rho}(t) - \mu_i(t)\| \leq C_d(1-\rho)$ for $i = 1, ..., d$, thus Condition 3.1.1 implies

$$\frac{\lambda_v(\mu(t))}{\lambda_v(\upsilon^{\rho}(t))} \leq \left(\frac{1}{\rho}\right)^{dC_1} \exp(f(1-\rho)). \tag{3.6.2}$$

Therefore, by the definition of L (3.1.3) and the fact that $\dot{\mu}(t) = \sum_{v\in\mathcal{V}} v\lambda_v(\mu(t))$, we have

$$\begin{aligned} L(\upsilon^{\rho}(t), \dot{\mu}(t)) &\leq \sum_{v\in\mathcal{V}} \lambda_v(\upsilon^{\rho}(t))\, \ell\left(\frac{\lambda_v(\mu(t))}{\lambda_v(\upsilon^{\rho}(t))}\right) \\ &\leq C_2 \log 1/\rho + C_2 f(1-\rho) + C_2 \end{aligned}$$

for some $C_2 < \infty$, where we apply (3.6.2) for the last inequality.

By (3.6.1) and the convexity and nonnegativity of $L(x,\cdot)$ stated in Proposition 3.5.2, one has

$$\begin{aligned} L\left(\upsilon^{\rho}(t),\dot{\upsilon}^{\rho}(t)\right) &\leq L\left(\upsilon^{\rho}(t),\dot{\gamma}(t)\right)+\rho L\left(\upsilon^{\rho}(t),\dot{\mu}(t)\right) \\ &\leq (1+c(\rho))\,L\left(\gamma(t),\dot{\gamma}(t)\right)+c_5(\rho)\,. \end{aligned}$$

for $c_5(\rho) = C_2\rho\log 1/\rho + C_3\rho f(1-\rho) + C_4\rho + c(\rho)$. Integrating both sides of the last inequality over $[0,1]$, we get

$$I\left(\upsilon^{\rho}\right) \leq (1+c(\rho))\,I(\gamma) + c_5(\rho)\,,$$

and thus iii) holds with $\upsilon = \upsilon^{\rho}$ for $\rho > 0$ sufficiently small. $\square$

In view of Lemma 3.6.2, it suffices to establish the lower bound for paths $\gamma \in AC([0,1]:\mathcal{S})$ with $I(\gamma) < \infty$ that satisfy the additional condition

$$\gamma_i(t) \geq b_0 t^D \text{ for some } b_0 > 0,\ D < \infty \text{ and for all } t \in [0,1]. \qquad (3.6.3)$$

3.6.2 Analysis for $t \in [0,\tau]$

Given $\delta > 0$, for $\tau > 0$ sufficiently small we can use excursion bounds for jump Markov processes (Lemma 3.6.4 below) to establish a lower bound for the quantity

$$\mathbb{P}\left(\sup_{t\in[0,\tau]} \|\mu^n(t) - \gamma(t)\| < \delta\right).$$

The more difficult part is to obtain for any $0 < \sigma < \delta$ a lower bound for $\mathbb{P}(\|\mu^n(\tau) - \gamma(\tau)\| < \sigma)$ that is uniform in $\mu^n(0)$ as long as $\|\mu^n(0) - \gamma(0)\|$ is sufficiently small.

Given $\varepsilon > 0$, $\tau \in (0,1]$, for any $\sigma > 0$, define the penalty function $g : \mathcal{S} \to \mathbb{R}$

$$g(x) = \begin{cases} 0 & ||x - \gamma(\tau)|| < \sigma, \\ 2\varepsilon & \text{else.} \end{cases} . \tag{3.6.4}$$

We then have

$$\mathbb{P}(\|\mu^n(\tau) - \gamma(\tau)\| < \sigma) + e^{-2n\varepsilon} \geq \mathbb{E}\left[\exp(-ng(\mu^n(\tau)))\right]. \tag{3.6.5}$$

Recall from Theorem 3.3.6

$$\begin{aligned} &-\frac{1}{n}\log \mathbb{E}\left[\exp(-ng(\mu^n(\tau)))\right] \\ &= \inf_{\bar{\alpha}\in\mathcal{A}_b^{\otimes|\mathcal{V}|}} \bar{\mathbb{E}}\left[\sum_{v\in\mathcal{V}}\int_0^\tau \lambda_v^n(\bar{\mu}^n(t))\,\ell\left(\frac{\bar{\alpha}_v(t)}{\lambda_v^n(\bar{\mu}^n(t))}\right)dt + g(\bar{\mu}^n(\tau)) : \bar{\mu}^n = \Lambda^n(\bar{\alpha},\mu^n(0))\right]. \end{aligned}$$

We now state the main result of this subsection.

Lemma 3.6.3. *Assume the sequence of deterministic initial conditions $\{\mu^n(0)\}_{n\in\mathbb{N}}$ converges to $\rho_0 \in \mathcal{S}$ as n tends to infinity. Then, given $\varepsilon > 0$ and $\delta > 0$, there exists $\tau > 0$ such that for any $\sigma > 0$, there exists $\eta = \eta(\sigma) > 0$, such that $\|\rho_0 - \gamma(0)\| \leq \eta$ implies*

$$\liminf_{n\to\infty}\frac{1}{n}\log \mathbb{P}\left(\|\mu^n(\tau) - \gamma(\tau)\| \leq \sigma, \sup_{0\leq s\leq\tau}\|\mu^n(s) - \gamma(s)\| \leq \delta\right) \geq -\frac{\varepsilon}{2}.$$

Proof. The idea is to argue that for large n and small τ, μ^n stays close to a communicating path (defined in Condition 3.1.4) that connects $\gamma(0)$ to $\gamma(\tau)$. Since the jump rate is bounded below away from zero along such a path, one obtains a nice upper bound for the cost. By Condition 3.1.4 and Remark 3.1.6, there exists a communicating path $\phi \in C([0,\tau] : \mathcal{S})$, with $\phi(0) = \gamma(0)$ and $\phi(\tau) = \gamma(\tau)$, and $F, U < \infty$, $\{v_m\}_{m=1}^F \subset \mathcal{V}$ and $0 = t_0 < t_1 < \cdots < t_F = \tau$, such that

$$\dot{\phi}(t) = \sum_{v \in \mathcal{V}} \bar{\alpha}_v(t) v, \text{ a.e. } t \in [0,\tau],$$

where

$$\bar{\alpha}_v(t) = \begin{cases} U 1_{[t_{m-1}, t_m)}(t) & \text{if } v = v_m, m = 1, ..., F \\ 0 & v \notin \{v_m\}_{m=1}^F. \end{cases}$$

Also, by Condition 3.1.4 and (3.6.3), there exist $c', p, D < \infty$ and $c_0, c > 0$, such that

$$\|\bar{\alpha}_v\|_\infty = \frac{1}{\tau}\int_0^\tau \left\|\dot{\phi}(t)\right\| dt \leq c' \frac{\|\gamma(\tau) - \gamma(0)\|}{\tau}, \tag{3.6.6}$$

and

$$\lambda_{v_m}(\phi(s)) \geq c\left(\min_i \gamma_i(\tau)\right)^p \geq c_0 \tau^{Dp}, \text{ if } s \in [t_{m-1}, t_m], \ m = 1, ..., F. \tag{3.6.7}$$

Define $\bar{\mu}^n = \Lambda^n(\bar{\alpha}, \mu^n(0))$, where Λ^n is as defined in (3.3.9). The LLN for Poisson random measures implies that $\{\bar{\mu}^n\}_{n\in\mathbb{N}}$ converges uniformly on $[0,\tau]$ in probability to $\bar{\mu}$, where $\bar{\mu}(0) = \rho_0$,

$$\frac{d}{dt}\bar{\mu}(t) = \sum_{v\in\mathcal{V}} \bar{\alpha}_v(t) v, \text{ a.e. } t \in [0,\tau]. \tag{3.6.8}$$

By the Lipschitz continuity of $\lambda_v(\cdot)$, the fact that $\|\bar{\mu}(s) - \phi(s)\| = \|\rho_0 - \gamma(0)\|$ since they use the same velocity, and (3.6.7), for any fixed τ, there exists some

$\eta_0(\tau) > 0$, such that for any $\eta \leq \eta_0(\tau)$, if $\|\rho_0 - \gamma(0)\| \leq \eta$, then

$$\lambda_{v_m}(\bar{\mu}(s)) \geq \frac{c_0}{2}\tau^{Dp}, \text{ for } s \in [t_{m-1}, t_m], \ m = 1, ..., F. \tag{3.6.9}$$

For $\eta \leq \eta_0(\tau)$, we can bound the cost along the path $\bar{\mu}$ by (each inequality will be explained below)

$$\begin{aligned} I_\tau(\bar{\mu}) &\leq \sum_{v\in\mathcal{V}} \int_0^\tau \lambda_v(\bar{\mu}(t))\,\ell\left(\frac{\bar{\alpha}_v(t)}{\lambda_v(\bar{\mu}(t))}\right) dt \\ &\leq |\mathcal{V}|\,\tau\left[c'\frac{\|\gamma(\tau)-\gamma(0)\|}{\tau}\left(\log\|\gamma(\tau)-\gamma(0)\| - \log\left(\frac{\frac{c_0}{2}\tau^{Dp+1}}{c'}\right)\right) + M\right] \\ &\leq C_4\|\gamma(\tau)-\gamma(0)\|\log\|\gamma(\tau)-\gamma(0)\| + C_4\|\gamma(\tau)-\gamma(0)\|\log\frac{1}{\tau} + C_4\tau \end{aligned} \tag{3.6.10}$$

for some $C_4 < \infty$, where M is defined in (2.3.1). Here we used the definition of L and (3.6.8) to obtain the first inequality; (3.6.6), (3.6.9), and (2.3.1) for the second inequality. By Lemma 3.5.6, $\|\gamma(\tau) - \gamma(0)\|\log\frac{1}{\tau} \to 0$ as $\tau \to 0$, together with the continuity of γ we obtain $I_\tau(\bar{\mu}) \to 0$.

We next bound the costs for the jump processes $\{\bar{\mu}^n\}_{n\in\mathbb{N}}$ by making use of its law of large numbers limit. For fixed τ, we use the fact that λ_v are Lipschitz continuous (by Condition 2.3.1), and $\lambda_{v_m}(\bar{\mu}(t))$ is uniformly bounded below away from zero if $\|\rho_0 - \gamma(0)\| \leq \eta_0(\tau)$ (by (3.6.9)). Therefore, $\lambda_v^n(\bar{\mu}^n(t))\,\ell(\bar{\alpha}_v/\lambda_v^n(\bar{\mu}^n(t)))$ converges in probability to $\lambda_v(\bar{\mu}(t))\,\ell(\bar{\alpha}_v/\lambda_v(\bar{\mu}(t)))$ uniformly for $t \in [0, \tau]$. Then for $\eta < \min\{\eta_0(\tau), \sigma/2\}$, applying the dominated convergence theorem and using the upper semicontinuity of g defined in (3.6.4), we have

$$\begin{aligned}
&\limsup_{n\to\infty} \bar{\mathbb{E}}\left[\sum_{v\in\mathcal{V}}\int_0^\tau \lambda_v^n\left(\bar{\mu}^n(t)\right)\ell\left(\frac{\bar{\alpha}_v(t)}{\lambda_v^n\left(\bar{\mu}^n(t)\right)}\right)dt + g(\bar{\mu}^n\left(\tau\right)) : \bar{\mu}^n = \Lambda^n\left(\bar{\alpha},\mu^n(0)\right)\right] \\
&\quad\le \bar{\mathbb{E}}\left[\sum_{v\in\mathcal{V}}\int_0^\tau \lambda_v\left(\bar{\mu}(t)\right)\ell\left(\frac{\bar{\alpha}_v(t)}{\lambda_v\left(\bar{\mu}(t)\right)}\right)dt + g(\bar{\mu}(\tau))\right] \\
&\quad\le C_4\left\|\gamma\left(\tau\right)-\gamma\left(0\right)\right\|\log\left\|\gamma\left(\tau\right)-\gamma\left(0\right)\right\| + C_4\left\|\gamma\left(\tau\right)-\gamma\left(0\right)\right\|\log\frac{1}{\tau} + C_4\tau,
\end{aligned}$$

where the last inequality follows from (3.6.10) and (3.6.4). Choose $\tau > 0$ sufficiently small such that the last expression is less than $\varepsilon/2$. Combining the last display with the representation formula, for all sufficiently large n and sufficiently small η, $\|\rho_0 - \gamma(0)\| < \eta$ implies

$$-\frac{1}{n}\log\mathbb{E}\left[\exp(-ng(\mu^n(\tau)))\right] \le \varepsilon/2. \tag{3.6.11}$$

When combined with (3.6.5), this gives a lower bound on $\mathbb{P}(\|\mu^n(\tau)-\gamma(\tau)\| < \sigma)$.

We will conclude the argument by establishing an upper bound for the excursion probability for μ^n during $[0,\tau]$. Given $\varepsilon > 0$, applying a standard martingale inequality (stated as Lemma 3.6.4 below), for sufficiently small τ we have

$$\mathbb{P}\left(\sup_{0\le s\le\tau}\|\mu^n(s)-\mu^n(0)\| > \frac{\delta}{3}\right) \le 2d\exp\left(-n\varepsilon\right).$$

On the other hand, by taking τ smaller if necessary we can guarantee that $\sup_{s\in[0,\tau]}\left\|\gamma\left(s\right)-\gamma\left(0\right)\right\| \le \delta/3$. It follows that for $\eta \in \left[0,\frac{\delta}{3}\right]$,

$$\begin{aligned}
\mathbb{P}\left(\sup_{0\le s\le\tau}\|\mu^n(s)-\gamma\left(s\right)\| > \delta\right) &\le \mathbb{P}\left(\sup_{0\le s\le\tau}\|\mu^n(s)-\mu^n(0)\| > \frac{\delta}{3}\right) \\
&\le 2d\exp\left(-n\varepsilon\right)
\end{aligned}$$

Combining this with estimates (3.6.5) and (3.6.11) we arrive at the desired conclusion. $\square$

The following lemma is an adaptation of Lemma 2.3 in [13]. The lemma follows from bounds for certain exponential martingales.

Lemma 3.6.4. *Let $C_1 = \max_{v\in\mathcal{V}} \|v\|, C_2 = M|\mathcal{V}|C_1$, and define*

$$\bar{\ell}(b) \doteq \frac{1}{C_1} b(\log(b/C_2) - 1)$$

for $b > C_2$. Then $\bar{\ell}(b)/b \to \infty$ as $b \to \infty$, and given any $\delta > 0$, for all $\tau \leq \frac{\delta}{2\sqrt{d}C_2}$

$$\mathbb{P}\left(\sup_{0\leq t\leq\tau} \|\mu^n(t) - \mu^n(0)\| \geq \delta\right) \leq 2d\exp\left(-\tau n\bar{\ell}\left(\frac{\delta}{2\sqrt{d}\tau}\right)\right).$$

3.6.3 Analysis for $t \in [\tau, 1]$

Let $B(x,r)$ denote the open Euclidean ball centered at x with radius r. Also, for $\gamma \in AC([\tau,1] : \mathcal{S})$ such that $\gamma(\tau) = y$, we denote

$$I^y(\gamma) = \int_\tau^1 L(\gamma(s), \dot{\gamma}(s))\, ds,$$

to emphasize the dependence on y. $\mathbb{P}_{y_n}$ and $\mathbb{E}_{y_n}$ denote the probability and expectation, respectively, conditioned on $\mu^n(\tau) = y_n$. Define the mapping $\Lambda^n_\tau : \mathcal{A}_b^{\otimes|\mathcal{V}|} \times \mathcal{S} \to D([\tau,1] : \mathcal{S})$ by

$$\Lambda^n_\tau(\bar{\alpha}, \rho)(t) = \rho + \sum_{v\in\mathcal{V}} v \int_{[\tau,t)} \int_{\mathcal{Y}} 1_{[0,\bar{\alpha}_v(s)]}(x) \frac{1}{n} N^n_v(dsdx),$$

for $\bar{\alpha} \in \mathcal{A}_b^{\otimes|\mathcal{V}|}$ and $\rho \in \mathcal{S}$. It suffices to study a path γ that satisfies (3.6.3). Therefore we can assume $I^y(\gamma) < \infty$, and that there exists $\xi > 0$ such that $\gamma(t)$ lies in $\mathcal{S}^\xi$ for $t \in [\tau, 1]$. We will prove the following uniform Laplace principle upper bound for $\{\mu^n(\cdot)\}_{n\in\mathbb{N}}$ on $[\tau, 1]$:

Proposition 3.6.5. *Let $\xi > 0$ as defined in the previous paragraph, and fix $y \in \mathcal{S}^\xi$. Then there exists $\sigma > 0$ such that for any bounded and Lipschitz continuous functional F on $D([\tau, 1] : \mathcal{S})$,*

$$\liminf_{n\to\infty} \inf_{y_n \in B(y,\sigma)} \left(\frac{1}{n} \log \mathbb{E}_{y_n} \left[\exp(-nF(\mu^n))\right] - G(y_n, F) \right) \geq 0, \tag{3.6.12}$$

where

$$G(y, F) = - \inf_{\gamma \in AC\left([\tau,1]:\mathcal{S}^\xi\right)} \left[I^y(\gamma) + F(\gamma)\right]. \tag{3.6.13}$$

In particular, this implies the uniform large deviation lower bound: for any $\varepsilon > 0$ and $\delta > 0$, there exists $\sigma > 0$ such that for any $y_n \in B(\gamma(\tau), \sigma)$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_{y_n} \left(\sup_{t\in[\tau,1]} \|\mu^n(t) - \gamma(t)\| < \delta \right) \geq -I^{\gamma(\tau)}(\gamma) - \frac{\varepsilon}{2}.$$

The proof of Proposition 3.6.5 relies on the following approximation argument, which we now establish.

Fix $y \in \mathcal{S}^\xi$ and a bounded and Lipschitz continuous functional F on $D([\tau, 1] : \mathcal{S})$. By Proposition 1.2.7 of [12], to prove (3.6.12), it suffices to show that for any sequence $\{y_n\}_{n\in\mathbb{N}}$ such that $\|y_n - y\| \to 0$ as $n \to \infty$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{E}_{y_n} \left[\exp(-nF(\mu^n))\right] \geq G(y, F). \tag{3.6.14}$$

In other words, it suffices to show that for any $\varepsilon > 0$ and $\gamma_\varepsilon \in AC\left([\tau, 1] : \mathcal{S}^\xi\right)$ such that $-\left(I^y(\gamma_\varepsilon) + F(\gamma_\varepsilon)\right) \geq G(y, F) - \varepsilon$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{E}_{y_n} \left[\exp(-nF(\mu^n))\right] \geq -\left(I^y(\gamma_\varepsilon) + F(\gamma_\varepsilon)\right),$$

or

$$\limsup_{n\to\infty} -\frac{1}{n} \log \mathbb{E}_{y_n} \left[\exp(-nF(\mu^n))\right] \leq I^y(\gamma_\varepsilon) + F(\gamma_\varepsilon).$$

Fix $\varepsilon > 0$ and denote γ_ε simply by γ. We now approximate γ by a piecewise linear path. Let $\Delta = \frac{1-\tau}{m}$ for some $m \in \mathbb{N}$. For $k = 0, 1, ..., m-1$ let $a_k^\Delta = \frac{1}{\Delta} \int_{\tau+k\Delta}^{\tau+(k+1)\Delta} \dot{\gamma}(s)\, ds$. Define

$$\dot{\gamma}^\Delta(t) = a_k^\Delta \quad \text{if } t \in (\tau + k\Delta, \tau + (k+1)\Delta),\ k = 0, ..., m-1,$$

and

$$\gamma^\Delta(t) = y + \int_\tau^t \dot{\gamma}^\Delta(s)\, ds \ \text{ for } t \in [\tau, 1]. \tag{3.6.15}$$

Then γ^Δ is the piecewise linear interpolation of the continuous process γ with mesh size Δ. Note that for any $v \in \mathcal{V}$ and $t \in [\tau, 1]$, $\lambda_v(\gamma^\Delta(t))$ is continuous and uniformly bounded away from zero. The proof of (3.6.12) thus relies on the following standard approximation result.

Lemma 3.6.6. *Let $\gamma \in AC\left([\tau, 1] : \mathcal{S}^\xi\right)$, and define γ^Δ as in (3.6.15). Then for any $\varepsilon > 0$, there exists $\Delta_0(\varepsilon) > 0$, such that for any $\Delta < \Delta_0(\varepsilon)$, and a.e. $t \in [\tau, 1]$, there exists a piecewise constant vector $q^\Delta(t)$ such that $\sum_{v\in\mathcal{V}} v q_v^\Delta(t) = \dot{\gamma}^\Delta(t)$, and*

$$\int_\tau^1 \sum_{v\in\mathcal{V}} \lambda_v(\gamma^\Delta(t)) \ell\left(\frac{q_v^\Delta(t)}{\lambda_v(\gamma^\Delta(t))}\right) dt \leq I^y(\gamma) + \varepsilon. \tag{3.6.16}$$

Proof. We first claim $\lim_{\Delta\to 0} I^y\left(\gamma^\Delta\right) = I^y(\gamma)$. Define $\left\{x_k^\Delta\right\}_{k=0}^{m-1} \in [0, \infty)^{|\mathcal{V}|}$ such that for $v \in \mathcal{V}$,

$$x_{k,v}^{\Delta} = \frac{1}{\Delta} \int_{\tau+k\Delta}^{\tau+(k+1)\Delta} \lambda_v (\gamma (s))\, ds.$$

Define $x^{\Delta}(t) = x_k^{\Delta}$ if $t \in [\tau + k\Delta, \tau + (k+1)\Delta)$. Recall the definition of $\bar{L}$ given in (3.5.4). By the joint convexity of $\bar{L}$ established in Lemma 3.5.2 and Jensen's inequality,

$$\begin{aligned} \int_{\tau}^{1} \bar{L}\left(x^{\Delta}(s), \dot{\gamma}^{\Delta}(s)\right) ds &= \Delta \sum_{k=0}^{m-1} \bar{L}\left(x_k^{\Delta}, a_k\right) \\ &\leq \sum_{k=0}^{m-1} \int_{\tau+k\Delta}^{\tau+(k+1)\Delta} \bar{L}\left(\lambda(\gamma(s)), \dot{\gamma}(s)\right) ds = I^y(\gamma). \end{aligned}$$

Since $I^y(\gamma) < \infty$ by the continuity of $\bar{L}$ and dominated convergence theorem we conclude that

$$\lim_{\Delta \to 0} \int_{\tau}^{1} \bar{L}\left(x^{\Delta}(s), \dot{\gamma}^{\Delta}(s)\right) ds = I^y(\gamma).$$

Furthermore, note that $\bar{L}\left(\lambda_v\left(\gamma^{\Delta}(s)\right), \dot{\gamma}^{\Delta}(s)\right) = L\left(\gamma^{\Delta}(s), \dot{\gamma}^{\Delta}(s)\right)$, and $\sup_{s\in[\tau,1]} \left\|\lambda_v\left(\gamma^{\Delta}(s)\right) - x^{\Delta}(s)\right\| \to 0$ as $\Delta \to 0$. Apply Lemma 3.5.7, we have

$$\begin{aligned} I^y\left(\gamma^{\Delta}\right) - \int_{\tau}^{1} \bar{L}\left(x^{\Delta}(s), \dot{\gamma}^{\Delta}(s)\right) ds &\leq c(\Delta) \sup_{\Delta>0} \int_{\tau}^{1} \bar{L}\left(x^{\Delta}(s), \dot{\gamma}^{\Delta}(s)\right) ds + c(\Delta) \\ &\leq c(\Delta)\, I^y(\gamma) + c(\Delta), \end{aligned}$$

for some $c(\Delta) \to 0$ as $\Delta \to 0$. The claim is proved on taking $\Delta \to 0$.

For $k = 0, ..., m-1$, fix some $t_k \in (\tau + k\Delta, \tau + (k+1)\Delta)$. We can take $q_k^{\Delta} \in [0, \infty)^{|\mathcal{V}|}$ such that $\sum_{v\in\mathcal{V}} v q_{k,v}^{\Delta} = a_k$, and

$$L(\gamma^{\Delta}(t_k), \dot{\gamma}^{\Delta}(t_k)) \geq \sum_{v\in\mathcal{V}} \lambda_v(\gamma^{\Delta}(t_k)) \ell\left(\frac{q_{k,v}^{\Delta}}{\lambda_v(\gamma^{\Delta}(t_k))}\right) - \frac{\varepsilon}{6}. \tag{3.6.17}$$

Since $\gamma^{\Delta} \in AC\left([\tau, 1] : \mathcal{S}^{\xi}\right)$, by the uniform continuity property of rate function in the state variable, which is stated in Proposition 3.5.5, choosing Δ sufficiently small we have

$$\left| I^{y}\left(\gamma^{\Delta}\right) - \sum_{k=0}^{m-1} \int_{k\Delta+\tau}^{(k+1)\Delta+\tau} L(\gamma^{\Delta}(t_k), \dot{\gamma}^{\Delta}(t_k))dt \right| < \frac{\varepsilon}{6}. \tag{3.6.18}$$

Define

$$q^{\Delta}(t) = q_k^{\Delta}(t), \text{ if } t \in (\tau + k\Delta, \tau + (k+1)\Delta), \; k = 0, ..., m-1.$$

Applying Lemma 3.5.7 again and integrate over $t \in [\tau, 1]$, there exists $c'(\Delta)$ which goes to zero as $\Delta \to 0$, such that

$$\begin{aligned} & \int_{\tau}^{1} \sum_{v\in\mathcal{V}} \lambda_v(\gamma^{\Delta}(t))\ell\left(\frac{q_v^{\Delta}(t)}{\lambda_v(\gamma^{\Delta}(t))}\right) dt & (3.6.19)\\ - & \sum_{k=0}^{m-1} \int_{k\Delta+\tau}^{(k+1)\Delta+\tau} \sum_{v\in\mathcal{V}} \lambda_v(\gamma^{\Delta}(t_k))\ell\left(\frac{q_{k,v}^{\Delta}}{\lambda_v(\gamma^{\Delta}(t_k))}\right) dt \\ \leq & \; c'(\Delta) \sum_{k=0}^{m-1} \int_{k\Delta+\tau}^{(k+1)\Delta+\tau} \sum_{v\in\mathcal{V}} \lambda_v(\gamma^{\Delta}(t_k))\ell\left(\frac{q_{k,v}^{\Delta}}{\lambda_v(\gamma^{\Delta}(t_k))}\right) dt + c'(\Delta). & (3.6.20) \end{aligned}$$

By (3.6.17), (3.6.18), and the fact that $I^{y}\left(\gamma^{\Delta}\right)$ is finite, one can take Δ smaller if necessary so that the LHS of (3.6.20) is less than $\varepsilon/6$. The conclusion follows by combining (3.6.17), (3.6.18), (3.6.20), and the convergence of $I^{y}\left(\gamma^{\Delta}\right)$ to $I^{y}(\gamma)$. $\square$

We now complete the proof of Proposition 3.6.5. By Lemma 3.6.6, for any $\varepsilon > 0$, there exists Δ sufficiently small and a collection of piecewise constant functions $\left\{q_v^{\Delta}(\cdot)\right\}_{v\in\mathcal{V}}$ on $[\tau, 1]$ that satisfy (3.6.16). It follows directly from the LLN for Poisson random measures that as $n \to \infty$, $\bar{\mu}^n = \Lambda_{\tau}^n\left(q^{\Delta}, y_n\right)$ converges uniformly on $[\tau, 1]$ in probability to γ^{Δ}. Therefore, by the uniform continuity of $\lambda_v(\cdot)\,\ell\left(q_v^{\Delta}/\lambda_v(\cdot)\right)$ on $\mathcal{S}^{\xi}$ and the uniform convergence of $\lambda_v^n(\cdot)$ to $\lambda_v(\cdot)$ by Condition 2.3.1, $\lambda_v^n(\bar{\mu}^n(\cdot))\,\ell\left(q_v^{\Delta}(\cdot)/\lambda_v^n(\bar{\mu}^n(\cdot))\right)$ converges uniformly on $[\tau, 1]$ in probability to $\lambda_v(\gamma^{\Delta}(\cdot))\ell\left(q_v^{\Delta}(\cdot)/\lambda_v(\gamma^{\Delta}(\cdot))\right)$. Combining the variational representation formula

(Theorem 3.3.6), (3.6.16), and the dominated convergence theorem, for any Lipschitz continuous functional F on $D([\tau,1]:\mathcal{S})$, we have

$$\begin{aligned}
&\limsup_{n\to\infty} -\frac{1}{n}\log \mathbb{E}_{y_n}\left[\exp(-nF(\mu^n))\right] \\
&\quad = \limsup_{n\to\infty} \inf_{\bar{\alpha}\in\mathcal{A}_b^{\otimes|\mathcal{V}|}} \bar{\mathbb{E}}_{y_n}\left[\begin{array}{c} \sum_{v\in\mathcal{V}}\int_\tau^1 \lambda_v^n\left(\bar{\mu}^n(t)\right)\ell\left(\frac{\bar{\alpha}_v(t)}{\lambda_v^n(\bar{\mu}^n(t))}\right)dt + F(\bar{\mu}^n): \\ \bar{\mu}^n = \Lambda_\tau^n\left(\bar{\alpha},y_n\right) \end{array}\right] \\
&\quad \le \limsup_{n\to\infty} \bar{\mathbb{E}}_{y_n}\left[\int_\tau^1 \sum_{v\in\mathcal{V}} \lambda_v^n\left(\bar{\mu}^n(t)\right)\ell\left(\frac{q_v^\Delta(t)}{\lambda_v^n\left(\bar{\mu}^n(t)\right)}\right)dt + F(\bar{\mu}^n): \bar{\mu}^n = \Lambda_\tau^n\left(q^\Delta,y_n\right)\right] \\
&\quad = \bar{\mathbb{E}}_y\left[\int_\tau^1 \sum_{v\in\mathcal{V}} \lambda_v(\gamma^\Delta(t))\ell\left(\frac{q_v^\Delta(t)}{\lambda_v(\gamma^\Delta(t))}\right)dt + F(\gamma^\Delta)\right] \\
&\quad \le \bar{\mathbb{E}}_y\left[I^y(\gamma) + \varepsilon + F(\gamma^\Delta)\right] \\
&\quad \le I^y(\gamma) + F(\gamma) + 2\varepsilon
\end{aligned}$$

for all Δ sufficiently small, where the last inequality follows from the continuity of F, and the fact that $\sup_{t\in[\tau,1]}\left\|\gamma^\Delta(t)-\gamma(t)\right\|\to 0$ as $\Delta\to 0$. By the Markov property, the LDP lower bound follows from Lemma 3.6.3 and Proposition 3.6.5.

3.7 The Locally Uniform LDP

We now turn to the proof of Theorem 3.1.15. Fix $t\in[0,1]$. As shown in Corollary 3.1.11, one can express the rate function J_t of $\{\mu^n(t)\}_{n\in\mathbb{N}}$ in terms of a variational problem. In what follows, fix $x\in\mathcal{S}$ and $\{x_n\}_{n\in\mathbb{N}}$ such that $x_n\in\mathcal{S}_n$ and $\|x_n-x\|\to 0$ as $n\to\infty$.

3.7.1 Proof of the upper bound

Given any $\varepsilon > 0$, recall that $B(x,\varepsilon)$ denote the Euclidean ball centered at x with radius ε. For n sufficiently large such that $x_n \in \bar{B}(x,\varepsilon)$, by the LDP upper bound stated in Corollary 3.1.11,

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\mu^n(t)=x_n\right) &\leq \limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\mu^n(t)\in\bar{B}(x,\varepsilon)\right)\\ &\leq -\bar{J}_t^{\varepsilon}(\mu_0,x),\end{aligned}$$

where we define

$$\bar{J}_t^{\varepsilon}(\mu_0,x)=\inf\left\{I_t(\gamma):\gamma\in D([0,1]:\mathcal{S}),\gamma(0)=\mu_0,\gamma(t)\in\bar{B}(x,\varepsilon)\right\}. \quad (3.7.1)$$

To prove the upper bound, it suffices to show that

$$\liminf_{\varepsilon\to 0}\bar{J}_t^{\varepsilon}(\mu_0,x)\geq J_t(\mu_0,x). \quad (3.7.2)$$

Lemma 3.7.1. *Assume Condition 3.1.14.i) holds. Then there exists a function* $c:[0,\infty)\to[0,\infty)$ *such that*

i). $c(\varepsilon)\to 0$ as $\varepsilon\to 0$, and

ii). given any $\varepsilon>0$ and $x,y\in\mathcal{S}$ such that $\|x-y\|<\varepsilon$, one can construct a path $\gamma\in AC([0,\varepsilon]:\mathcal{S})$ such that $\gamma(0)=x$, $\gamma(\varepsilon)=y$, and $J_\varepsilon(x,y)\leq I_\varepsilon(\gamma)\leq c(\varepsilon)$, where $I_\varepsilon, J_\varepsilon$ are defined in (3.1.5), (3.1.9), respectively.

Before proving Lemma 3.7.1, we first describe how it can be used to prove Lemma 3.1.16 and (3.7.2).

Proof of Lemma 3.1.16. For any $\varepsilon > 0$, take $t > 0$ and $\gamma \in AC([0,t] : \mathcal{S})$ such that $\gamma(0) = x$, $\gamma(t) = y$, and $I_t(\gamma) \le V(x,y) + \varepsilon/2$. Given $\delta > 0$, and any $y^\delta \in \mathcal{S}$ such that $\left\|y^\delta - y\right\| \le \delta$, by Lemma 3.7.1, there exists a path $\nu \in AC([0,\delta] : \mathcal{S})$ with $\nu(0) = y$, $\nu(\delta) = y^\delta$ with $I_\delta(\nu) \le c(\delta)$. Let $\bar{\gamma}$ be the concatenation of γ and ν. Then we have

$$V\left(x, y^\delta\right) \le I_{t+\delta}(\bar{\gamma}) = I_t(\gamma) + I_\delta(\nu) \le V(x,y) + \varepsilon/2 + c(\delta).$$

It suffices to choose δ such that $c(\delta) \le \varepsilon/2$. The other inequality (and the joint continuity with respect to both variables) are proved in the same way. $\square$

We can complete the proof of the locally uniform LDP upper bound as follows. For $\delta > 0$, pick $\gamma \in AC([0,1] : \mathcal{S})$ such that $\gamma(0) = \mu_0, \gamma(t) \in \bar{B}(x,\varepsilon)$, and $I_t(\gamma) \le \bar{J}_t^\varepsilon(\mu_0, x) + \delta$. By Lemma 3.7.1 there exists a path $\nu \in AC([0,\varepsilon] : \mathcal{S})$ with $\nu(0) = \gamma(t)$, $\nu(\varepsilon) = x$ with $I_\varepsilon(\nu) \le c(\varepsilon)$, where $c(\varepsilon) \to 0$ as $\varepsilon \to 0$. Let $\bar{\gamma}$ be the concatenation of γ and ν. We now rescale $\overline{\gamma}$ to obtain a new path: for $c = (t+\varepsilon)/t$, define $\bar{\gamma}_c \in AC([0,t] : \mathcal{S})$ by $\bar{\gamma}_c(s) = \bar{\gamma}(cs)$, $s \in [0,t]$. Then $\bar{\gamma}_c(0) = \mu_0$, $\bar{\gamma}_c(t) = x$. Moreover, by Proposition 3.5.8, for ε sufficiently small, $I_t(\bar{\gamma}_c) \le I_{t+\varepsilon}(\overline{\gamma}) + \delta$, and by the construction above,

$$J_t(\mu_0, x) \le I_t(\bar{\gamma}_c) \le I_{t+\varepsilon}(\overline{\gamma}) + \delta = I_t(\gamma) + I_\varepsilon(\nu) + \delta \le \bar{J}_t^\varepsilon(\mu_0, x) + 2\delta + c(\varepsilon).$$

Taking the limit inferior as $\varepsilon \to 0$ and then sending $\delta \to 0$, (3.7.2) follows.

Proof of Lemma 3.7.1. By Condition 3.1.14.i) and Remark 3.1.6, there exists a strongly communicating path $\gamma \in AC([0,\varepsilon] : \mathcal{S})$ such that $\gamma(0) = x$, $\gamma(\varepsilon) = y$, with constant speed $U \le c' \|x - y\| / \varepsilon \le c'$. Precisely, there exist $F < \infty$ and

$0 = t_0 < t_1 < \cdots < t_F = 1$, such that

$$\dot{\gamma}(t) = \sum_{m=1}^{F} U v_m 1_{[t_{m-1}\varepsilon, t_m\varepsilon)}(t) \text{ for a.e. } t \in [0, \varepsilon].$$

Since $I_\varepsilon(\gamma) = \sum_{m=1}^{F} \left(I_{t_m\varepsilon}(\gamma) - I_{t_{m-1}\varepsilon}(\gamma) \right)$, it suffices to bound each term from above.

Recall from (3.1.11) that for any $j \in \mathcal{N}_m$, $\langle e_j, v_m \rangle < 0$. Let $b_1 \doteq \min_{m=1,\ldots,F} \min_{j\in\mathcal{N}_m} |\langle e_j, v_m\rangle| > 0$. Note that for $s \in [t_{m-1}\varepsilon, t_m\varepsilon)$, and any $j \in \mathcal{N}_m$, $\gamma_j(t_m\varepsilon) - \gamma_j(s) = -\langle e_j, v_m\rangle U (t_m\varepsilon - s)$, and thus $\gamma_j(s) \geq b_1 U (t_m\varepsilon - s)$. Therefore, by Definition 3.1.12, there exist constants $c_1 > 0$, $p_1 < \infty$, such that for ε sufficiently small,

$$\lambda_{v_m}(\gamma(s)) \geq c_1 \left(\prod_{j\in\mathcal{N}_m} \gamma_j(s) \right)^{p_1} \geq \tilde{c}_1 U^\kappa (t_m\varepsilon - s)^\kappa,$$

where $\kappa = dp_1 < \infty$ and $\tilde{c}_1 = c_1 b_1^{dp_1} > 0$. Thus, by taking $q_{v_m} = U$, and $q_v = 0$ for $v \neq v_m$ in the first line below, we have

$$\begin{aligned} L(\gamma(s), \dot{\gamma}(s)) &= \inf_{q:\sum_{v\in\mathcal{V}} vq_v = \dot{\gamma}(s)} \sum_{v\in\mathcal{V}} \lambda_v(\gamma(s))\, \ell\left(\frac{q_v}{\lambda_v(\gamma(s))} \right) \\ &\leq \lambda_{v_m}(\gamma(s))\, \ell\left(\frac{U}{\lambda_{v_m}(\gamma(s))} \right) + \sum_{v\in\mathcal{V}} \lambda_v(\gamma(s)) \\ &\leq U \log\left(\frac{U}{\tilde{c}_1 U^\kappa (t_m\varepsilon - s)^\kappa} \right) - U + C_2 \\ &\leq -(\kappa - 1) U \log U - \kappa U \log(t_m\varepsilon - s) - U(1 + \log \tilde{c}_1) + C_2, \end{aligned}$$

for some constant $C_2 < \infty$. Therefore,

$$\begin{aligned} I_{t_m\varepsilon}(\gamma) - I_{t_{m-1}\varepsilon}(\gamma) &= \int_{t_{m-1}\varepsilon}^{t_m\varepsilon} L(\gamma(s), \dot{\gamma}(s))\, ds \\ &\leq -C_3(U)\, \varepsilon \log\varepsilon + C_4(U)\, \varepsilon \end{aligned}$$

for some constants $C_3(U), C_4(U)$ such that $\sup_{U\in[0,c']}(C_3(U)\vee C_4(U))<\infty$. Summing over m, we have $J_\varepsilon(\mu_0,y)\le I_\varepsilon(\gamma)\le c(\varepsilon)$, where $c(\varepsilon)=O(\varepsilon|\log\varepsilon|)$ as $\varepsilon\to 0$. □

3.7.2 Proof of the Lower Bound

For the proof of the lower bound, take any $\varepsilon>0$ small. Then by the Markov property for $\{\mu^n\}$, we have

$$\mathbb{P}_{\mu_0}(\mu^n(t)=x_n)\ge \mathbb{P}_{\mu_0}(\mu^n(t-\varepsilon)\in B(x,\varepsilon))\cdot \inf_{w_n\in B(x,\varepsilon)\cap\mathcal{S}_n}\mathbb{P}_{w_n}(\mu^n(\varepsilon)=x_n).$$

The LDP lower bound in Corollary 3.1.11 implies

$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}_{\mu_0}(\mu^n(t-\varepsilon)\in B(x,\varepsilon))\ge -J^\varepsilon_{t-\varepsilon}(\mu_0,x),$$

where

$$J^\varepsilon_t(\mu_0,x)\doteq\inf\{I_t(\gamma):\gamma\in D([0,1]:\mathcal{S}),\gamma(0)=\mu_0,\gamma(t)\in B(x,\varepsilon)\}.$$

The proof of the lower bound will be complete if we can show both of the following:

i) $\limsup_{\varepsilon\to 0}J^\varepsilon_{t-\varepsilon}(\mu_0,x)\le J_t(\mu_0,x)$.

ii) The **Local Communication Property**: There exist a function $c:[0,\infty)\to[0,\infty)$ that satisfies $c(\varepsilon)\to 0$ as $\varepsilon\to 0$ and is such that for all $\varepsilon>0$ sufficiently small,

$$\inf_{w_n\in B(x,\varepsilon)\cap\mathcal{S}_n}\mathbb{P}_{w_n}(\mu^n(\varepsilon)=x_n)\ge\exp(-nc(\varepsilon)+o(n)).$$

To prove the first property, we will use Proposition 3.5.8. For any $\delta > 0$, take $\gamma \in AC([0,1] : \mathcal{S})$ such that $\gamma(t) = x$ and $I_t(\gamma) \leq J_t(\mu_0, x) + \delta$. Take $c = t/(t-\varepsilon)$ and consider the path $\gamma_c \in AC([0, t-\varepsilon] : \mathcal{S})$, such that $\gamma_c(s) = \gamma(cs)$, $s \in [0, t]$. Then $\gamma_c(0) = \mu_0$, $\gamma_c(t-\varepsilon) = x$. By Proposition 3.5.8, given $\delta > 0$, for ε sufficiently small, $I_{t/c}(\gamma_c) \leq I_t(\gamma) + \delta$, and we have

$$J^{\varepsilon}_{t-\varepsilon}(\mu_0, x) \leq I_{t-\varepsilon}(\gamma_c) = I_{t/c}(\gamma_c) \leq I_t(\gamma) + \delta \leq J_t(\mu_0, x) + 2\delta,$$

and the conclusion follows by taking first $\varepsilon \to 0$ and then $\delta \to 0$.

Proof of local communication property. We will use Condition 3.1.14 and Lemma 3.6.1. Fix some $w_n \in B(x, \varepsilon) \cap \mathcal{S}_n$, note that the probability of $\mu^n(\varepsilon) = x_n$ is no less than the probability that μ^n hitting x_n at ε by passing through a given discrete strongly communicating path ϕ_n that connects w_n and x_n.

By Condition 3.1.14, there exists $F < \infty$, $0 = t_0 \leq t_1 \leq \cdots \leq t_F = 1$, $\{v_m\}_{m=1}^{F}$, and constants $c_1 > 0$, $c', p_1 < \infty$, such that the image of ϕ_n is $\mathcal{S}_n \cap \{\phi(s) : s \in [0, \varepsilon]\}$, for some $\phi \in AC([0, \varepsilon] : \mathcal{S})$ that satisfies $\phi(0) = x$, and

$$\dot{\phi}(s) = \sum_{m=1}^{F} U v_m 1_{[t_{m-1}\varepsilon, t_m\varepsilon)}(s), \quad \text{a.e. } s \in [0, \varepsilon],$$

with $U = c' \|x_n - w_n\| / \varepsilon \leq c'$. Also, for $s \in [t_{m-1}\varepsilon, t_m\varepsilon)$ and large n, $\lambda^n_{v_m}(\phi_n(s)) > c_1(\prod_{j \in \mathcal{N}_m} (\phi_n)_j(s))^{p_1}$. Let $z^{(m)} = \phi_n(t_m\varepsilon-)$. By the Markov property $\mathbb{P}_{w_n}(\mu^n(\varepsilon) = x_n) \geq \prod_m \mathbb{P}_{z^{(m)}}\left(\mu^n((t_{m+1} - t_m)\varepsilon) = z^{(m+1)}\right)$, and it suffices to give a lower bound for each term in the product. This will be proved by comparison with another Markov process Z^n. Thus, without modifying the notation, we let $\mu^n(t)$ denote the process stopped when it first leaves the set of points $\{\phi_n(s) : s \in [t_{m-1}\varepsilon, t_m\varepsilon)\}$. For each m and $t \in [0, (t_{m+1} - t_m)\varepsilon)$, define Z^n to be the jump Markov process with

$Z^n(0) = z^{(m)}$, with the same set $\mathcal{V}$ of jump directions, and jump rates

$$\bar{\lambda}_v(x) = \begin{cases} nc_1 \left(\prod_{j \in \mathcal{N}_m} x_j \right)^{p_1} & \text{if } v = v_m \\ nM & \text{if } v \in \mathcal{V}/v_m \end{cases}$$

so long as Z^n stays in the set $\{\phi_n(s) : s \in [t_{m-1}\varepsilon, t_m\varepsilon)\}$, and with the process stopped when it jumps off the line segment. Note that $\bar{\lambda}_v(x)$ bounds $\lambda^n_{v_m}(x)$ from below in the set, while nM is an upper bound on all jump rates.

Let $p(t)$ be the probability distribution of $Z^n(t)$: for any $x \in \mathcal{S}_n$, $p_x(t) = \mathbb{P}(Z^n(t) = x)$. It satisfies the Kolmogorov forward equation

$$\dot{p}(t) = \sum_{x,y \in \mathcal{S}_n} (e_y - e_x) \bar{A}_{xy}, \tag{3.7.3}$$

where

$$\bar{A}_{xy} = \begin{cases} \bar{\lambda}_{n(y-x)}(x) & \text{if } n(y-x) \in \mathcal{V} \\ 0 & \text{else.} \end{cases}$$

Also, the distribution of $\mu^n(t)$ satisfies

$$\dot{m}(t) = \sum_{x,y \in \mathcal{S}_n} (e_y - e_x) A_{xy},$$

where

$$A_{xy} = \begin{cases} \lambda_{n(y-x)}(x) & \text{if } n(y-x) \in \mathcal{V} \\ 0 & \text{else.} \end{cases}$$

By construction

$$\mathbb{P}_{x_m}(\mu^n((t_{m+1} - t_m)\varepsilon) = x_{m+1}) \geq \mathbb{P}_{x_m}(Z^n((t_{m+1} - t_m)\varepsilon) = x_{m+1}).$$

If l is the number of points in the discrete segment $\{\phi_n(s) : s \in [t_m\varepsilon, t_{m+1}\varepsilon)\}$, then by Definition 3.1.13 $l \leq C_2 n\varepsilon$ for some $C_2 < \infty$. The product of the jump rates of Z^n along this segment satisfies

$$\prod_{x:x\in\{\phi_n(s):s\in[t_m\varepsilon,t_{m+1}\varepsilon)\}} \left(nc_1\left(\prod_{j\in\mathcal{N}_m} x_j\right)^{p_1}\right) \geq c_1^l\,(l!)^{k_m p_1}/n^{(k_m p_1-1)l},$$

where $k_m \doteq |\mathcal{N}_m| \leq d$. The lower bound in the last inequality is achieved when $\{\phi_n(s) : s \in [t_m\varepsilon, t_{m+1}\varepsilon)\}$ is a segmemt that ends up on $x_{m+1} \in \partial\mathcal{S}$, and for all $j \in \mathcal{N}_m$, $x_j = 1, ..., l$ along the segment. Then it follows from Lemma 3.6.1 that for $\varepsilon > 0$ sufficiently small,

$$\begin{aligned}
&\mathbb{P}_{x_m}(\mu^n((t_{m+1}-t_m)\varepsilon) = x_{m+1})\\
&\quad\geq \mathbb{P}_{x_m}(z^n((t_{m+1}-t_m)\varepsilon) = x_{m+1})\\
&\quad\geq \frac{1}{l!}c_1^l\,(l!)^{k_m p_1}/n^{(k_m p_1-1)l}\,((t_{m+1}-t_m)\varepsilon)^l \exp(-nM|\mathcal{V}|(t_{m+1}-t_m)\varepsilon)\\
&\quad\geq c_1^l\left(\frac{l!}{n^l}\right)^{dp_1-1}\varepsilon^l\exp(-nM|\mathcal{V}|\varepsilon)\\
&\quad\geq \exp(nC_2\varepsilon\log c + (dp_1-1)\,nC_2\varepsilon\log(C_2\varepsilon/e) + nC_2\varepsilon\log\varepsilon - nM|\mathcal{V}|\varepsilon + o(\varepsilon n))\\
&\quad= \exp(dp_1 n(C_2\varepsilon\log\varepsilon + O(\varepsilon)) + o(n)),
\end{aligned}$$

where for the third inequality we used the fact that the function $x \mapsto x^l\exp(-nbx)$ is decreasing for l/n sufficiently small, and for the fourth inequality we used Stirling's approximation.

From this, we conclude $\mathbb{P}_w(\mu^n(\varepsilon) = x_n) \geq \exp(-nc(\varepsilon) + O(\varepsilon) + o(n))$ with $c(\varepsilon) = O(\varepsilon\log\varepsilon)$, as desired. $\square$

Chapter Four

Markov Chain Approximation for Quasipotentials

This chapter studies Markov chain approximations for the quasipotential function associated with the stationary measure of the mean field interacting particle systems. In Section 4.1 we define the quasipotential as an optimal control problem, and state basic assumptions that will be used in the proofs in Sections 4.2. Section 4.2 contains the proof of a comparison lemma, which allows us to approximate the quasipotential in a small region by solving a linear quadratic regulator problem. In Section 4.3 we present the Markov chain approximation algorithm, following the general framework of [29]. In Section 4.4 we construct different approximation schemes, and discuss the method to solve the associated optimization problem. Finally, some numerical examples are shown in Section 4.5.

4.1 Assumptions

We focus on the mean field interacting particle systems described in Chapter 2 with $K = 1$. Specifically, the empirical measure process $\{\mu^n(t), t \geq 0\}$ is a càdlàg jump Markov process that takes values in $\mathcal{S}_n$, with its generator given by (2.2.1). We will denote $\Gamma^1_{ij}(\cdot)$ and $\alpha^1_{ij}(\cdot)$ (which are defined in Condition 2.3.3) simply as $\Gamma_{ij}(\cdot)$ and $\alpha_{ij}(\cdot)$.

Following Theorem 2.3.2, $\{\mu^n\}$ satisfies a functional law of large numbers limit μ, that solves the Kolmogorov forward equation

$$\begin{cases} \dot{\mu}(t) = \mu(t)^T \Gamma(\mu(t)), \\ \mu(0) = \mu_0. \end{cases} \tag{4.1.1}$$

And by Theorem 3.1.10, $\{\mu^n\}$ satisfy the sample path large deviation principle (LDP) with local rate function

$$L(x,\beta) = \inf_{q:\mathbf{1}^T q=\beta} \sum_{i=1}^{d} \sum_{j=1,j\neq i}^{d} \alpha_{ij}(x)\,\ell\left(\frac{q_{ij}}{\alpha_{ij}(x)}\right), \tag{4.1.2}$$

for $x \in \mathcal{S}$ and $\beta \in \Delta^{d-1} \doteq \{\beta \in \mathbb{R}^d : \sum_{i=1}^{d} \beta_i = 0\}$. Here q is a $d \times d$ matrix with nonnegative off-diagonal entries and $q_{ii} = -\sum_{j=1,j\neq i}^{d} q_{ij}$, and $\mathbf{1} = (1, ..., 1)$ is a d-dimensional vector.

Fix $x_0 \in \mathcal{S}$, and let V be the quasipotential associated with $\{x_0\}$. Specifically, we define

$$\begin{aligned} V(x) &= \inf\left\{ \begin{array}{c} \int_0^\tau L\left(\phi(s), \dot{\phi}(s)\right) ds : \phi \in AC([0,\tau] : \mathcal{S}), \\ \phi(0) = x_0, \phi(\tau) = x, \tau < \infty \end{array} \right\} \\ &= \inf\left\{ \begin{array}{c} \int_0^\tau L\left(\phi(s), -\dot{\phi}(s)\right) ds : \phi \in AC([0,\tau] : \mathcal{S}), \\ \phi(0) = x, \phi(\tau) = x_0, \tau < \infty \end{array} \right\}. \end{aligned} \tag{4.1.3}$$

We call L the running cost of the associated control problem. It is equivalent to take the infimum in (4.1.3) over all $\phi \in AC([0,\infty) : \mathcal{S})$ such that $\lim_{t\to\infty} \phi(t) = x_0$, see [10], Lemma 2.

By Lemma 3.1.16, the quasipotential function thus defined is continuous (and thus uniformly bounded) in $\mathcal{S}$. We also denote

$$V \doteq \sup_{x\in\mathcal{S}} V(x). \tag{4.1.4}$$

Let H be the Hamiltonian which is defined to be the Legendre transform of L:

for $x \in \mathcal{S}$ and $p \in \Delta^{d-1}$,

$$\begin{aligned} H(x,p) &\doteq \sup_{\beta \in \Delta^{d-1}} \{\langle \beta, p \rangle - L(x,\beta)\} \\ &= \sum_{i=1}^{d} \sum_{j=1, j\neq i}^{d} \alpha_{ij}(x)\left(\exp(p_j - p_i) - 1\right). \end{aligned} \tag{4.1.5}$$

The following result follows from Lemma 3.5.2.

Proposition 4.1.1. *For any $x \in \mathcal{S}$ fixed, both $L(x,\cdot)$ and $H(x,\cdot)$ are strictly convex on Δ^{d-1}.*

For simplicity we will use the following reduced coordinates. The coordinate transformations $\mathcal{I}_1 : (x_1, ..., x_d) \to (x_1, ..., x_{d-1})$ and $\mathcal{I}_2 : (x_1, ..., x_d) \to (x_2 - x_1, ..., x_d - x_1)$ maps $\mathcal{S}$ and Δ^{d-1} respectively, into $\bar{\mathcal{S}} = \{x \in \mathbb{R}^{d-1} : x_i \geq 0 \text{ and } \sum_{i=1}^{d-1} x_i \leq 1\}$ and $\mathbb{R}^{d-1}$. Under such coordinate transformations, for $i = 1, ..., d-1$, one can denote $\tilde{x}_i = x_i$, $\tilde{p}_i = p_{i+1} - p_1$. Then $\tilde{x}, \tilde{p} \in \mathbb{R}^{d-1}$, and we can write H as

$$H(\tilde{x},\tilde{p}) = \sum_{i=1}^{d} \sum_{j=1, j\neq i}^{d} \alpha_{ij}(\tilde{x})\left(\exp(\tilde{p}_{j-1} - \tilde{p}_{i-1}) - 1\right), \tag{4.1.6}$$

where $\tilde{p}_0 \doteq 0$. With some abuse of notation, we still denote $\tilde{x}, \tilde{p}$ as x, p respectively, in later sections.

Denote $\text{int}(\mathcal{S})$ as the relative interior of $\mathcal{S}$, and note that $\mathcal{I}_1(\text{int}(\mathcal{S}))$ is open in $\mathbb{R}^{d-1}$. For any open set $D \subset \mathbb{R}^m$, $m \leq d$, and $k \in \mathbb{N}$, denote $C^k(D : \mathbb{R})$ as the space of k times continuously differentiable functions on D. We will sometimes abuse the notation, and write $f \in C^k(\text{int}(\mathcal{S}) : \mathbb{R})$ and $g \in C^k\left(\Delta^{d-1} : \mathbb{R}\right)$ if $f \circ \mathcal{I}_1^{-1} \in$

$C^k(\mathcal{I}(\mathrm{int}(\mathcal{S})) : \mathbb{R})$ and $g \circ \mathcal{I}_2^{-1} \in C^k\left(\mathbb{R}^{d-1} : \mathbb{R}\right)$. For $f \in C^1\left(\mathrm{int}(\mathcal{S}) \times \Delta^{d-1} : \mathbb{R}\right)$, we use $D_{x_i} f$ and $D_{p_i} f$ to denote the partial derivative of $f \circ \mathcal{I}_1^{-1}$ with respect to $\tilde{x}_i$ and the partial derivative of $f \circ \mathcal{I}_2^{-1}$ with respect to $\tilde{p}_i$. For $k \in \mathbb{N}$, higher order derivatives D_x^k and D_p^k are defined in the same way.

The equilibrium point x_0 of the law of large numbers dynamics (4.1.1) satisfies $L(x_0, 0) = 0$. In other words,

$$\sum_{i \neq j} \alpha_{ij}(x_0) - \sum_{k \neq j} \alpha_{jk}(x_0) = 0,\ j = 1, ..., d.$$

A direct computation gives

$$D_{p_j} H(x_0, 0) = \sum_{i \neq j+1} \alpha_{i,j+1}(x_0) - \sum_{k \neq j+1} \alpha_{j+1,k}(x_0) = 0,\ j = 1, ..., d-1, \qquad (4.1.7)$$

Condition 4.1.2. *Assume that there is a unique unstable equilibrium point $x_0 \in$int$(\mathcal{S})$ of the law of large numbers dynamics (4.1.1), in the sense that all eigenvalues of the Hessian $D_x D_p H(x_0, 0)$ has negative real parts.*

In this chapter we focus on the quasipotential associated with the unique unstable equilibrium $\{x_0\}$. We also assume some regularity condition on $\alpha_{ij}(\cdot)$.

Condition 4.1.3. *Assume $\alpha_{ij}(\cdot) \in C^3(int(\mathcal{S}) : \mathbb{R})$.*

Remark 4.1.4. *Condition 4.1.3 implies that H, and therefore its Legendre dual L, are C^3 on int$(\mathcal{S}) \times \Delta^{d-1}$, in both variables. For H, the C^3-regularity is a direct consequence from (4.1.6). For L, this follows from the fact that since $L(x, \cdot)$ is strictly convex on the closed set Δ^{d-1} (by Proposition 4.1.1), the supremum in (4.1.5) is uniquely attained, and thus L has the same regularity as H (a similar argument is carried out in Lemma 4.2.1 below).*

Applying Theorem 2 of [10], Condition 4.1.3 implies that the quasipotential V defined in (4.1.3) is C^3 in some open neighborhood (relative to $\mathcal{S}$) of x_0 (which we denote as D). We now define the region of strong regularity of V.

Definition 4.1.5. *For V defined as in (4.1.3), a relatively open set $\Omega_0 \subset \mathcal{S}$ is called a region of strong regularity of V if*

i). $V \in C^3 (\Omega_0 : [0, \infty))$.

ii). For any initial condition $x \in \Omega_0$, there exists a unique the optimal trajectory ϕ^* in (4.1.3), that satisfies $V(x) = \int_0^\infty L\left(\phi^*(s), -\dot{\phi}^*(s)\right) ds$, and the image $\{\phi^*(s) : s \in [0, \infty)\} \subset \Omega_0$.

For general definitions and properties of region of strong regularity, we refer the reader to [19], Section 6.7. As shown there in the case of small noise diffusion problems, the relatively open set Ω_0 is connected, contains some neighborhood of x_0, and is also dense in $\mathcal{S}$.

The optimal control problem (4.1.3) that defines the quasipotential V can be associated with the solution of a Hamilton-Jacobi-Bellman equation. V can be characterized (cf. [10], Section 3) as the maximal viscosity subsolution to the PDE

$$\begin{aligned} H(x, DV(x)) &= 0, \text{ for } x \in \mathcal{S} \setminus \{x_0\} \\ V(x_0) &= 0, \end{aligned} \tag{4.1.8}$$

where $DV \doteq \left(D_{x_1}V, ..., D_{x_{d-1}}V\right)$, and H is understand as in (4.1.6). Restricted to Ω_0, V is a classical (C^3) solution to (4.1.8) with a local minimum at x_0 (Corollary 5 of [10]).

4.2 A Comparison Lemma

In this section we provide a comparison lemma for two quasipotentials. We will construct an approximation of the quasipotential (4.1.3), denoted as $V^{(2)}$, in a neighborhood of x_0, by taking the approximated running cost associated with the same exit problem. The approximate quasipotential can be solved explicitly using linear quadratic regulators. Later in Section 4.3 we will construct a discrete approximation of the solution to (4.1.8), and in particular its boundary value at x_0 will be approximated by $V^{(2)}$ in the intersection of the lattice and a ball contains x_0. In what follows we assume all the conditions stated in Section 4.1.

By Condition 4.1.3 and (4.1.6), we see that $H(x,0)=0$, and

$$D_x^k H(x,0) = 0,\ k=1,2 \tag{4.2.1}$$

for any $x \in$int$(\mathcal{S})$. Thus, recalling (4.1.7), for (x,p) near $(x_0,0)$ we have the following expansion:

$$\begin{aligned} H(x,p) &= \frac{1}{2}p^T Ap + \frac{1}{2}(x-x_0)^T Bp + \frac{1}{2}p^T B^T (x-x_0) + O\left(\|x-x_0\|^3 + \|p\|^3\right) \\ &= \frac{1}{2}p^T Ap + (x-x_0)^T Bp + O\left(\|x-x_0\|^3 + \|p\|^3\right), \end{aligned} \tag{4.2.2}$$

where, using (4.1.6),

$$\begin{aligned} A_{ij} &= D_{p_i} D_{p_j} H(x_0,0) \\ &= \begin{cases} \sum_{k=1,k\neq i+1}^{d} \alpha_{k,i+1}(x_0) + \sum_{k=1,k\neq i+1}^{d} \alpha_{i+1,k}(x_0) & \text{if } i=j, \\ -\alpha_{i+1,j+1}(x_0) - \alpha_{j+1,i+1}(x_0) & \text{if } i\neq j, \end{cases} \end{aligned} \tag{4.2.3}$$

and

$$\begin{aligned} B_{ij} &= D_{x_i} D_{p_j} H(x_0, 0) \\ &= \sum_{k=1, k\neq j+1}^{d} \frac{\partial}{\partial x_i} \alpha_{k,j+1}(x_0) - \sum_{k=1, k\neq j+1}^{d} \frac{\partial}{\partial x_i} \alpha_{j+1,k}(x_0). \end{aligned} \tag{4.2.4}$$

Since $H(x, \cdot)$ is strictly convex on Δ^{d-1} by Proposition 4.1.1, A is a positive definite matrix. Define

$$H^{(2)}(x, p) \doteq \frac{1}{2} p^T A p + (x - x_0)^T B p.$$

The Legendre transform of $H^{(2)}$ takes the form

$$\begin{aligned} L^{(2)}(x, \beta) &= \sup_{p \in \Delta^{d-1}} \left[\langle \beta, p \rangle - \left(\frac{1}{2} p^T A p + (x - x_0)^T B p \right) \right] \\ &= \frac{1}{2} \left(\beta - B^T (x - x_0) \right)^T A^{-1} \left(\beta - B^T (x - x_0) \right), \end{aligned} \tag{4.2.5}$$

for $x \in \mathcal{S}$ and $\beta \in \Delta^{d-1}$. We claim that $L^{(2)}$ is the true quadratic approximation of L, in the following sense.

Lemma 4.2.1. *There exists $\varepsilon > 0$, such that for any $(x, \beta) \in \mathcal{S} \times \Delta^{d-1}$ with $\|x - x_0\| < \varepsilon$ and $\|\beta\| < \varepsilon$, $L(x, \beta) = L^{(2)}(x, \beta) + O\left(\|x - x_0\|^3 + \|p\|^3\right)$.*

Proof. By taking ε smaller if necessary, we can assume without loss of generality that $x \in$int$(\mathcal{S})$. By Remark 4.1.4, $L \in C^3(\text{int}(\mathcal{S}) \times \Delta^{d-1} : \mathbb{R})$. It suffices to show that the partial derivatives of L and $L^{(2)}$ at $(x_0, 0)$ are equal up to the second order. It is clear from the definition of the equilibrium point that $L(x_0, 0) = L^{(2)}(x_0, 0) = 0$. Since $L(x, \beta) = \sup_p (\langle \beta, p \rangle - H(x, p))$, the fact that H is strictly convex in p (Proposition 4.1.1) implies that the supremum is attained at some unique $p^* = p^*(x, \beta)$. By the implicit function theorem, the strict convexity of $H(x,\cdot)$ and the fact that H is C^3 in both variables (Remark 4.1.4), p^* is C^3 in both variables, satisfies $p^*(x_0,0)=0$ and

$$D_p H(x, p^*(x,\beta)) = \beta, \text{ for } (x,\beta) \in int(\mathcal{S}) \times \Delta^{d-1}.$$

Using the fact that

$$L(x,\beta) = \langle \beta, p^*(x,\beta)\rangle - H(x, p^*(x,\beta))$$

and $DH(x_0,0)=0$, we obtain

$$\begin{aligned}
D_x L(x_0,0) &= D_p L(x_0,0) = 0,\\
D^2_{xx} L(x_0,0) &= -D^2_{xx} H(x_0,0),\\
D^2_{x\beta} L(x_0,0) &= -D^2_{xp} H(x_0,0)\left(D^2_{pp} H(x_0,0)\right)^{-1},\\
D^2_{\beta\beta} L(x_0,0) &= \left(D^2_{pp} H(x_0,0)\right)^{-1}.
\end{aligned}$$

And similarly,

$$\begin{aligned}
D_x L^{(2)}(x_0,0) &= D_p L^{(2)}(x_0,0) = 0,\\
D^2_{xx} L^{(2)}(x_0,0) &= -D^2_{xx} H^{(2)}(x_0,0),\\
D^2_{x\beta} L^{(2)}(x_0,0) &= -D^2_{xp} H^{(2)}(x_0,0)\left(D^2_{pp} H^{(2)}(x_0,0)\right)^{-1},\\
D^2_{\beta\beta} L^{(2)}(x_0,0) &= \left(D^2_{pp} H^{(2)}(x_0,0)\right)^{-1}.
\end{aligned}$$

The claim then follows by noting that the partial derivatives of H and $H^{(2)}$ at $(x_0,0)$ are equal up to the second order. $\square$

Note that $V(x_0)=0$. Since $H(x_0,0)=0$, $H(x_0,\cdot)$ is nonnegative and strictly convex, $H(x_0,p)=0$ if and only if $p=0$. Therefore within a neighborhood of x_0, V

is C^3 and $DV(x_0) = 0$ by (4.1.8). One can then expand V near x_0 as

$$V(x) = (x - x_0)^T \bar{P}(x - x_0) + O\left(\|x - x_0\|^3\right). \tag{4.2.6}$$

Since $(x_0, 0)$ is an unstable equilibrium for the corresponding Hamiltonian system, $\bar{P}$ is a symmetric positive definite matrix (see Section 4 of [10]). We are interested in obtaining the second order approximation of V near x_0, i.e. in identifying $\bar{P}$ in (4.2.6). Define $V^{(2)}$ as the solution to the same exit problem as in (4.1.3) but with the running cost $L^{(2)}$ instead of L:

$$V^{(2)}(x) = \inf \left\{ \begin{array}{c} \int_0^\tau L^{(2)}\left(\phi(s), -\dot{\phi}(s)\right) ds : \phi \in AC([0,\tau] : \mathcal{S}), \\ \phi(0) = x, \phi(\tau) = x_0, \tau < \infty \end{array} \right\}. \tag{4.2.7}$$

Substituting the expression for $L^{(2)}$ from (4.2.5), we can rewrite $V^{(2)}$ as

$$V^{(2)}(x) = \inf \left\{ \begin{array}{c} \int_0^\tau \frac{1}{2} u(s)^T A^{-1} u(s)\, ds : \dot{\phi} = -B^T(\phi - x_0) - u, \\ \phi(0) = x, \phi(\tau) = x_0, \tau < \infty \end{array} \right\}. \tag{4.2.8}$$

This is the control problem associated with the linear quadratic regulator, which is known to admit an explicit solution. Specifically, it is known (see e.g. Chapter 8.2 of [37]) to be a quadratic function $V^{(2)}(x) = (x - x_0)^T P(x - x_0)$, $x \in \mathcal{S}$, with P being the maximal solution of the algebraic Ricatti equation,

$$BP + PB^T + 2PAP = 0, \tag{4.2.9}$$

where A and B are defined in (4.2.3) and (4.2.4), respectively.

In the next theorem, we show that $P = \bar{P}$, and $V^{(2)}$ is indeed the second order approximation of V.

Theorem 4.2.2. *There exists $\delta > 0$, such that for any $\|x - x_0\| < \delta$, we have*

$$\left| V(x) - V^{(2)}(x) \right| = O\left(\|x - x_0\|^3 \right).$$

Proof. Since $DV(x) = 2\bar{P}(x - x_0) + O(\|x - x_0\|^2)$, substitute the expansion of Hamiltonian (4.2.2) into the Hamilton-Jacobi equation (4.1.8), we obtain

$$(x - x_0)^T \left(B\bar{P} + \bar{P}B^T + 2\bar{P}A\bar{P} \right)(x - x_0) + O\left(\|x - x_0\|^3 \right) = 0,$$

and therefore

$$B\bar{P} + \bar{P}B^T + 2\bar{P}A\bar{P} = 0. \tag{4.2.10}$$

The maximality of P in (4.2.9) therefore implies $V(x) \leq V^{(2)}(x) + O(\|x - x_0\|^3)$.

To prove the other inequality, for y in some neighborhood of x_0, (4.1.8) and the definition of H via (4.1.5) implies the optimal velocity $\beta^* = \beta^*(y)$ is uniquely attained, and satisfies

$$\langle DV(y), \beta^* \rangle - L(y, \beta^*) = 0. \tag{4.2.11}$$

And the optimal path φ of (4.1.3) takes the form

$$\begin{aligned} \dot{\varphi}(t) &= -\beta^*(\varphi(t)), \qquad (4.2.12) \\ \varphi(0) &= x. \end{aligned}$$

By the implicit function theorem for vector valued functions, $\beta^*(\cdot)$ is at least C^2 in a neighborhood of x_0, $\beta^*(x_0) = 0$, and near x_0 it admits Taylor expansion $\beta^*(y) = E(y - x_0) + O(\|y - x_0\|^2)$. Now recall the quadratic approximation of L near $(x_0, 0)$

in (4.2.5). Looking at the $O(\|x-x_0\|^2)$ term in (4.2.11), one obtains

$$E^T\bar{P}+\bar{P}E-\frac{1}{2}\left(E^T-B\right)A^{-1}\left(E-B^T\right)=0.$$

Using (4.2.10), we see that $E=B^T+2A\bar{P}$ solves the above algebraic equation. To see the solution is unique, take $E=B^T+2A\bar{P}+X$ in the above equation, an we obtain $\frac{1}{2}XA^{-1}X=0$, the positive definiteness of A thus implies $X=0$. Therefore,

$$\beta^*(y)=\left(B^T+2A\bar{P}\right)(y-x_0)+O(\|y-x_0\|^2).$$

Note that $U(x)=(x-x_0)^T\bar{P}(x-x_0)$ is a Lyapunov function to (4.2.12). In fact, there exists $\delta, c_0>0$, such that for any $\|\varphi(t)-x_0\|<\delta$,

$$\begin{aligned}\frac{d}{dt}U(\varphi(t)) &= (\varphi(t)-x_0)^T\left(-B\bar{P}-\bar{P}B^T-4\bar{P}A\bar{P}\right)(\varphi(t)-x_0)\\ &+ O\left(\|\varphi(t)-x_0\|^3\right)\\ &= -2(\varphi(t)-x_0)^T\bar{P}A\bar{P}(\varphi(t)-x_0)+O\left(\|\varphi(t)-x_0\|^3\right)\\ &\leq -c_0\|\varphi(t)-x_0\|^2.\end{aligned}$$

Where we use (4.2.10) to obtain the second equality, and the strict positive definiteness of $\bar{P}A\bar{P}$ to obtain the last inequality. Also, the strict postitive definiteness of $\bar{P}$ implies there exists $K_1>0$ and $K_2<\infty$ such that $K_1\|y-x_0\|^2\leq U(y)\leq K_2\|y-x_0\|^2$. Applying a version of Lyapunov exponential stability theorem for quadratic Lyaponov functions (see, e.g. Theorem 3.4 of [32]), there exist $C_1=C_1(\delta), C_2=C_2(\delta)>0$, $C_3=C_3(\delta)$, such that for $\|x-x_0\|<\delta$, $\|\varphi(t)-x_0\|\leq C_1\|x-x_0\|e^{-C_2t}$ and $\|\dot{\varphi}(t)\|\leq C_3\|x-x_0\|e^{-C_2t}$. Taking this solution to be the control picked in (4.2.8), we obtain constants $C_4, C_5<\infty$, such that

$$
\begin{aligned}
V^{(2)}(x) &\leq \int_0^{\infty} L^{(2)}(\varphi(s), -\dot{\varphi}(s))\, ds \\
&\leq \int_0^{\infty} \left[L(\varphi(s), -\dot{\varphi}(s)) + C_4 \left(\|\varphi(t) - x_0\|^3 + \|\dot{\varphi}(t)\|^3 \right) \right] ds \\
&\leq V(x) + C_5 \|x - x_0\|^3 \int_0^{\infty} e^{-3C_2 t} dt \\
&= V(x) + \frac{C_5}{3C_2} \|x - x_0\|^3 .
\end{aligned}
$$

□

4.3 Markov Chain Approximations

We now adapt the methods in [29] and [4] to construct a discrete approximation to the solution of the optimal control problem (4.1.3). Following [4], we construct value functions defined on lattice approximations of $\mathcal{S}$, that solve discrete analogs of Hamilton-Jacobi-Bellman equations. Before describing the construction we point out the special features of the model that did not appear in [29] and [4]. First, the set of all possible jump directions of the Markov process (2.2.1) is $\mathcal{V} \doteq \{e_j - e_i : i, j = 1, ..., d, i \neq j\}$, which is different from the usual nearest neighbor random walks on square lattices. Second, the boundary for our exit problem is the singleton $\{x_0\}$. We take a sequence of sets $B^h \subset \mathcal{S}^h$ (that shrink to $\{x_0\}$ as $h \to 0$) to approximate this boundary, and assign the boundary conditions on B^h by the quadratic approximation $V^{(2)}$ studied in the previous section (which can be obtained by solving (4.2.9) for its Hessian). These differences will affect our construction of the Markov chain approximations and the convergence proof.

We now construct a discrete time, discrete state controlled random walk that approximates the dynamics $-\dot{\phi} = u$ as $h \to 0$.

For $h > 0$, we consider the controlled Markov chain ξ^h with state space $\mathcal{S}^h \doteq h\mathbb{Z}^d \cap \mathcal{S}$. Since by taking an affine map $\mathcal{S}$ maps to some $\bar{\mathcal{S}} \subset \mathbb{R}^{d-1}$, $\mathcal{S}^h$ can also be identified with $h\mathbb{T}^{d-1} \cap \bar{\mathcal{S}}$, where $\mathbb{T}^{d-1}$ is the $d-1$ dimensional triangular (simplicial) lattice. Also, define $B^h = \left\{x \in \mathcal{S}^h : \|x - x_0\| \leq C_0 h^{1/3}\right\}$, where $C_0 < \infty$ is a constant to be chosen, and since it will not affect the convergence rate of the algorithm, we set $C_0 = 1$ for simplicity.

The control space and dynamics can be described as follows. In the limiting calculus of variation problem (4.1.3), the control takes value in measurable functions on $[0, \infty)$. In discrete time formulation of the Markov chain approximation, for fixed $h > 0$, we work with time and state dependent controls $\{u_j^h(x) : j \in \mathbb{N}, x \in \mathcal{S}^h\}$, with $u_j^h(\cdot)$ taking value in the control space Δ^{d-1}. One can construct the controlled random walk using different class of controls. We choose to work with feedback controls, because it is simple, also because the state space $\mathcal{S}^h$ is finite, and the running cost L is strictly convex (thus grows superlinearly) in the control, the optimal control sequence for our discrete value function (introduced in (4.3.6) below) is uniquely attained.

Given the feedback control $\{u_j^h(\cdot)\}_{j \in \mathbb{N}}$, we can define a discrete time controlled random walk ξ^h. Namely, for $j \in \mathbb{N}$, ξ_{j+1}^h is obtained by updating ξ_j^h with transition probability $p^h(\xi_j^h, y | u_j^h(\xi_j^h))$, the time interpolation $\Delta t^h(\xi_j^h, u_j^h(\xi_j^h))$. According to [29], p^h and Δt^h should be chosen to satisfy the local consistency relations:

$$h \sum_{v \in \mathcal{V}} v p^h(x, x + hv | u) = u \Delta t^h(x, u), \tag{4.3.1}$$

and

$$\sum_{v\in\mathcal{V}}\left\|hv-u\Delta t^{h}\left(x,u\right)\right\|^{2}p^{h}\left(x,x+hv|u\right)=o\left(\|u\|\,\Delta t^{h}\left(x,u\right)\right). \tag{4.3.2}$$

Take a subset $\mathcal{V}^{*}\subset\mathcal{V}$, so that Δ^{d-1} can be decomposed as union of cones $\mathcal{C}_i$, each of which is generated by $d-1$ linearly independent vectors in $\mathcal{V}^{*}$, which we denote as $\mathcal{B}_i$:

$$\mathcal{C}_i=\left\{u\in\Delta^{d-1}:u=\sum_{v\in\mathcal{B}_i}c_v v,\,c_v\geq 0\right\}.$$

We also require these cones are minimal, in the sense that for any $v\in\mathcal{V}^{*}\backslash\mathcal{B}_i$, $v\notin\mathcal{C}_i$. Given u belongs to some $\mathcal{C}_i$, u can be written uniquely as a linear combination

$$u=\sum_{v\in\mathcal{B}_i}u_v v \tag{4.3.3}$$

for non-negative $\{u_v\}_{v\in\mathcal{B}_i}$, and we set $u_v=0$ for $v\in\mathcal{V}\backslash\mathcal{B}_i$. One natural choice of $\left(p^h,\Delta t^h\right)$ is to set

$$p^{h}\left(x,y|u\right)=\begin{cases}\frac{u_v}{\|u\|_1} & \text{if } y=x+hv,\,v\in\mathcal{B}_i\\ 0 & \text{otherwise}\end{cases},\ \text{for } u\in\mathcal{C}_i, \tag{4.3.4}$$

where $\|u\|_1\doteq\sum_{v\in\mathcal{V}}u_v$, and

$$\Delta t^{h}\left(u\right)=\frac{h}{\|u\|_1}. \tag{4.3.5}$$

Let $D\left([0,\infty):\mathbb{R}^d\right)$ denote the space of càdlàg paths in $\mathbb{R}^d$. Equipped with the Skorohod metric, this becomes a complete separable metric space. One can also take a continuous time interpolation, so that $\{\xi_j^h\}_{j\in\mathbb{N}},\{u_j^h(\cdot)\}_{j\in\mathbb{N}}$ become elements of $D\left([0,\infty):\mathbb{R}^d\right)$. They can be constructed recursively as follows. At $t=0$, take $\xi^h(0)=\xi_0^h\in\mathcal{S}^h$, $u^h(0)=u_0^h\left(\xi_0^h\right)\in\Delta^{d-1}$, and set $\Delta t^h\left(u_0^h\right)$ by (4.3.5). For $s\in\left[0,\Delta t^h\left(u_0^h\right)\right)$, define $\xi^h(s)=\xi_0^h$ and $u^h(s)=u_0^h\left(\xi_0^h\right)$. Suppose for some $k\in\mathbb{N}$,

$\left\{\xi_i^h\right\}_{i\le k-1}$ and $\left\{u_i^h\right\}_{i\le k-1}$ have been defined, we set $t_k = \sum_{i=0}^{k-1} \Delta t^h \left(u_i^h\left(\xi_i^h\right)\right)$. At t_k, update the value of ξ^h (denote as ξ_k^h) using (4.3.4) with the control $u_{k-1}^h\left(\xi_{k-1}^h\right)$, and denote $u^h\left(t_k\right) \doteq u_k^h\left(\xi_k^h\right)$. Let $t_{k+1} = \sum_{i=0}^{k} \Delta t^h\left(u_i^h\left(\xi_i^h\right)\right)$, and for $s\in[t_k, t_{k+1})$, define $\xi^h(s) = \xi_k^h$ and $u^h(s) = u_k^h\left(\xi_k^h\right)$. We will switch between the discrete and continuous time description of ξ^h and u^h, whichever is more convenient.

Given a feedback control $\{u_j^h(\cdot)\}_{j\in\mathbb{N}}$ and a controlled random walk $\{\xi_j^h\}_{j\in\mathbb{N}}$, we can decompose the random walk into a drift part and a local martingale with mean zero: $\xi_k^h = \mathbb{E}\left(\xi_k^h | u_{k-1}^h\left(\xi_{k-1}^h\right)\right) + m_k^h$, where $\mathbb{E}\left(\xi_k^h | u_{k-1}^h\left(\xi_{k-1}^h\right)\right) = h u_{k-1}^h / \left\|u_{k-1}^h\right\|_1$.

Now consider the following optimal control problem on $\mathcal{S}^h$. For $x\in B^h$, we set $V^h(x) = V^{(2)}(x)$. For $x\in \mathcal{S}^h\backslash B^h$, the value function is defined by

$$V^h(x) = \inf_{u^h\in D\left([0,\infty):\Delta^{d-1}\right)} \mathbb{E}_x\left[\int_0^{\tau^h} L(\xi^h(s), -u^h(s))ds + V^{(2)}(\xi^h(\tau^h))\right], \qquad (4.3.6)$$

where $\mathbb{E}_x$ denotes the expectation conditioned on $\xi^h(0) = x$, and the exit time $\tau^h = \inf\left\{s : \xi^h(s)\in B^h\right\}$.

It is shown in [4], Section 3 (see also Chapter 5.8 of [29]), that V^h solves the discrete dynamic programming equation

$$\begin{aligned} V^h(x) &= \inf_{u\in\Delta^{d-1}}\left[\sum_{v\in\mathcal{V}} p^h(x, x+hv|u) V^h(x+hv) + L(x,-u)\Delta t^h(u)\right], x\in\mathcal{S}^h\backslash B^h, \\ V^h(x) &= V^{(2)}(x), x\in B^h, \end{aligned} \qquad (4.3.7)$$

where p^h and Δt^h are given by (4.3.4) and (4.3.5), respectively. This is our starting point for numerical approximations. Note that the discrete dynamic programming equation also holds when x is on $\partial\mathcal{S}$.

V^h can also be interpreted as a solution to a discrete HJB equation. For $u \in \Delta^{d-1}$, let $\{u_v\}_{v\in\mathcal{V}}$ be given by (4.3.3), and we define D_u^h to be the weighted finite difference operator, given by

$$D_u^h f(x) = \sum_{v\in\mathcal{V}} u_v \frac{f(x+hv) - f(x)}{h}, \; x \in \mathcal{S}^h,$$

for every bounded function $f : \mathcal{S}^h \to \mathbb{R}$. Then, subtracting $V^h(x)$ from both sides in (4.3.7), and dividing by $\Delta t^h(u)$, one obtains

$$\begin{aligned} 0 &= \inf_{u\in\Delta^{d-1}} \left[D_u^h V^h(x) + L(x, -u) \right], \; x \in \mathcal{S}^h \backslash B^h, \\ V^h(x) &= V^{(2)}(x), \; x \in B^h. \end{aligned} \tag{4.3.8}$$

4.4 Numerical Approximations

We now construct numerical schemes for the Markov chain approximation that satisfy local consistency properties (4.3.1) and (4.3.2). The controlled random walk should be defined in such a way that the transition probability has a simple expression, the dynamic programming equation (4.3.7) can be explicitly solved, and the data can propagate quickly from the boundary. We discuss below two choices of controlled random walks that satisfy these criteria.

Note that in the problem (4.3.7) there are two layers of infimization: the first infimization is over all the controlled jump rates (in the definition of $L(x, -u)$ in (4.1.2)) that lead to a given drift vector u, the second infimization is over $u \in \Delta^{d-1}$. We start with a lemma that simplifies the problem into a single infimization over the controlled rates.

Lemma 4.4.1. *For any $h > 0$ and $x \in \mathcal{S}^h$,*

$$\begin{aligned}
&\inf_{u \in \Delta^{d-1}} \left[\sum_{v \in \mathcal{V}} p^h\left(x, x + hv | u\right) V^h\left(x + hv\right) + L\left(x, -u\right) \Delta t^h\left(u\right) \right] &(4.4.1)\\
= &\inf_{\{\bar{a}_v\} \in [0,\infty)^{|\mathcal{V}|}} \left\{ \sum_{v \in \mathcal{V}} p^h\left(x, x + hv | - \sum_{v \in \mathcal{V}} v \bar{a}_v\right) V^h\left(x + hv\right) \right.\\
&\left. + \left(\sum_{v \in \mathcal{V}} \alpha_v\left(x\right) \ell\left(\frac{\bar{a}_v}{\alpha_v\left(x\right)}\right)\right) \Delta t^h\left(-\sum_{v \in \mathcal{V}} v \bar{a}_v\right) \right\}. &(4.4.2)
\end{aligned}$$

Proof. Recall that $L\left(x, -u\right) = \inf_{\{\bar{a}_v\}: -u = \sum v \bar{a}_v} \sum_{v \in \mathcal{V}} \alpha_v\left(x\right) \ell\left(\frac{\bar{a}_v}{\alpha_v(x)}\right)$. For any fixed $u \in \Delta^{d-1}$, since the function $\{\bar{a}_v\} \mapsto \sum_{v \in \mathcal{V}} \alpha_v\left(x\right) \ell\left(\frac{\bar{a}_v}{\alpha_v(x)}\right)$ is strictly convex on $[0, \infty)^{|\mathcal{V}|}$, the infimum in the definition of $L\left(x, -u\right)$ is uniquely attained at some $\{\bar{a}_v^*\left(x, u\right)\} \in [0, \infty)^{|\mathcal{V}|}$ such that $-u = \sum_{v \in \mathcal{V}} v \bar{a}_v^*$. Therefore,

$$\begin{aligned}
&\sum_{v \in \mathcal{V}} p^h\left(x, x + hv | u\right) V^h\left(x + hv\right) + L\left(x, -u\right) \Delta t^h\left(u\right)\\
= &\sum_{v \in \mathcal{V}} p^h\left(x, x + hv | u\right) V^h\left(x + hv\right) + \left(\sum_{v \in \mathcal{V}} \alpha_v\left(x\right) \ell\left(\frac{\bar{a}_v^*}{\alpha_v\left(x\right)}\right)\right) \Delta t^h\left(u\right).
\end{aligned}$$

Taking the infimum over $u \in \Delta^{d-1}$ one obtains the left hand side of (4.4.1) is greater than or equal to the right hand side.

To prove the other inequality, take any $\{\bar{a}_v\} \in [0, \infty)^{|\mathcal{V}|}$, and let $u \in \Delta^{d-1}$ be given by $-u = \sum_{v \in \mathcal{V}} v \bar{a}_v$. Then we have

$$\begin{aligned}
&\sum_{v \in \mathcal{V}} p^h\left(x, x + hv | u\right) V^h\left(x + hv\right) + L\left(x, -u\right) \Delta t^h\left(u\right)\\
\leq &\sum_{v \in \mathcal{V}} p^h\left(x, x + hv | u\right) V^h\left(x + hv\right) + \left(\sum_{v \in \mathcal{V}} \alpha_v\left(x\right) \ell\left(\frac{\bar{a}_v}{\alpha_v\left(x\right)}\right)\right) \Delta t^h\left(u\right).
\end{aligned}$$

Infimize over all $\{\bar{a}_v\} \in [0, \infty)^{|\mathcal{V}|}$, one obtains the reverse inequality. $\square$

We next state a result which gives the structure of the minimum of the optimization problem (4.4.1). The strictly convexity and uniform superlinear growth of $L(x,\cdot)$ implies the local minimum of (4.4.1) is uniquely attained, and is also the global minimum. We start with a result in convex analysis proved in [4].

Lemma 4.4.2 (Lemma 7.1 of [4])**.** *Let $g:[0,\infty)\to\mathbb{R}$ be strictly convex and satisfy $g(0)>0$ and $\lim_{k\to\infty} g(z)/z=\infty$. Then there is a unique point $x\in(0,\infty)$ that achieves the infimum in $\inf_{z\geq 0} g(z)/z$. x is also the unique local minimum of the map $z\mapsto g(z)/z$.*

Lemma 4.4.3. *There exists a unique local minimizer of the optimization problem (4.4.1), which is also the global minimizer.*

Proof. For $z\geq 0$, define

$$\begin{aligned} g(z) &= \inf_{u\in\Delta^{d-1}:\|u\|_1=z} z\left[\sum_{v\in\mathcal{V}} p^h(x,x+hv|u)\,V^h(x+hv)+L(x,-u)\,\Delta t^h(u)\right] \\ &= \inf_{u\in\Delta^{d-1}:\|u\|_1=z}\left[\sum_{v\in\mathcal{V}} u_v V^h(x+hv)+L(x,-u)\,h\right], \end{aligned}$$

where we used the definition 4.3.4 and 4.3.5 to obtain the last equality. By Proposition 4.1.1, we see that g is the minimum of a strictly convex function subject to an affine constraint, which is itself strictly convex. Moreover, for fixed z, there exists a unique $u^*\in\Delta^{d-1}$, $\|u^*\|_1=z$ such that the minimum is attained. For $x\notin B^h$, $g(0)=L(x,0)\,h>0$. The superlinear growth of L (see Proposition 3.5.4) implies that $\lim_{k\to\infty} g(z)/z=\infty$. The conclusion then follows by applying Lemma 4.4.2. $\square$

4.4.1 Iterative solver: The widest neighborhood structure

We employ Gauss-Seidel iterative method to solve the dynamical programming equation (4.3.7) (see Section 4.5 for more details). By Lemma 4.4.3, if one can find a local minimizer to (4.3.7), it gives the global minimum. The strategy to solve the minimizer in (4.3.7) or (4.4.2) is to solve the infimization in each cone $\mathcal{C}_i$ (introduced in Section 4.3), and then minimize over all cones. To put it in more detail, the first step is to consider the relative interior of a cone, and look for local minimum as if there were no constraints. We will do this computation in Section 4.4.1 and 4.4.2. If the infimum of this problem is finite, some nonnegativity constraints are tested for each candidate minimizer. If these constraints are satisfied, then a local minimum is found. If none of the candidate minimizer satisfies the constraints, then the unique minimizer must be on the boundary of one of the cones. We then search these boundaries that form lower dimensional $(d-2, d-3, \ldots)$ hyperplanes of $\mathbb{R}^d$. The procedure ends by searching through all the lower dimensional boundaries until the local minimum is found.

We now focus on the optimization problem (4.4.2) inside a given cone. Different choices of $\mathcal{V}^*$ lead to different $\{\mathcal{C}_i\}$ (which we call "neighborhood structures"), and the choice should balance the tradeoff between the explicit solvability of the infimizer in (4.4.2), and the number of constrained minimization problems that must be solved as dimension grows. We present below two different neighborhood structures in Section 4.4.1 and 4.4.2.

We start with a neighborhood structure with the widest possible cone, so that boundary data can propagate quickly through the iteration. The disadvantage of this approach, as we will see below, is that the infimization problem becomes more complicated as d gets large.

Take a set of vectors $(w_1, ..., w_d) = (e_2 - e_1, e_3 - e_2, ..., e_1 - e_d)$, so that any $d-1$ of them form a basis for Δ^{d-1}. For any $x \in \mathcal{S}^h \backslash B^h$, take $\mathcal{V}^* \doteq \{w_i : i = 1, ..., d\}$. Since $\mathcal{V}^* \subset \mathcal{V}$, for each $v \in \mathcal{V}^*$, we use $\bar{a}_{ij}$ to denote $\bar{a}_v$ in (4.4.2) if $v = e_j - e_i$. Given $u \in \Delta^{d-1}$, we can write

$$\begin{aligned} -u &= \sum_{i=1}^{d} \sum_{j=1, j\neq i}^{d} \bar{a}_{ij} \left(e_j - e_i\right) \\ &= \sum_{j=1}^{d} \sum_{i=1}^{j-1} \bar{a}_{ij} \left(\sum_{k=i}^{j-1} w_k \right) + \sum_{j=1}^{d} \sum_{i=j+1}^{d} \bar{a}_{ij} \left(- \sum_{k=j}^{i-1} w_k \right) \\ &= \sum_{k=1}^{d-1} \left(\sum_{i=1}^{k} \sum_{j=k+1}^{d} \bar{a}_{ij} - \sum_{i=k+1}^{d} \sum_{j=1}^{k} \bar{a}_{ij} \right) w_k. \end{aligned}$$

Suppose u belongs to the positive cone spanned by $\mathcal{B}_1 \doteq \{w_1, ..., w_{d-1}\}$, the other cases can be treated similarly. Then although the choice of $\{\bar{a}_{ij}\}$ in the equation above is not unique, $\sum_{i=1}^{k} \sum_{j=k+1}^{d} \bar{a}_{ij} - \sum_{i=k+1}^{d} \sum_{j=1}^{k} \bar{a}_{ij}$ is uniquely determined for $k = 1, ..., d-1$. Recalling $\|u\|_1 = \sum_{v \in \mathcal{V}} u_v$, we see that

$$\|u\|_1 = - \sum_{k=1}^{d-1} \left(\sum_{i=1}^{k} \sum_{j=k+1}^{d} \bar{a}_{ij} - \sum_{i=k+1}^{d} \sum_{j=1}^{k} \bar{a}_{ij} \right) = \sum_{i=1}^{d} \sum_{j=1, j\neq i}^{d} (i-j)\, \bar{a}_{ij}. \qquad (4.4.3)$$

We construct the controlled Markov chain such that for $k = 1, ..., d-1$, take

$$p^h\left(x, x + hw_k | u\right) = \frac{\sum_{i=k+1}^{d} \sum_{j=1}^{k} \bar{a}_{ij} - \sum_{i=1}^{k} \sum_{j=k+1}^{d} \bar{a}_{ij}}{\|u\|_1}, \text{ if } x + hw_k \in \mathcal{S}^h$$

and $p^h\left(x, y | u\right) = 0$ otherwise. Also, set

$$\Delta t^h\left(u\right) = \frac{h}{\|u\|_1}.$$

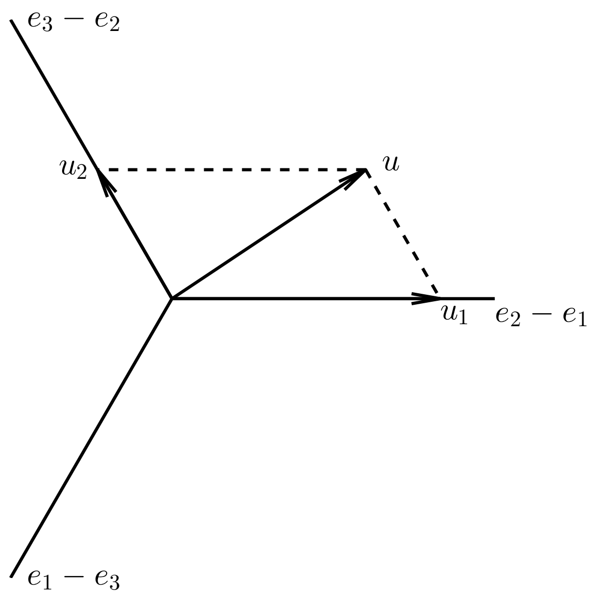

Figure 4.1: Controlled random walks using the widest neighborhood structure for $d = 3$

One can check such a choice of p^h and Δt^h satisfies (4.3.1) and (4.3.2). We then solve the following unconstrained minimization problem

$$\inf_{\{\bar{a}_{ij}\}} \left\{ \sum_{k=1}^{d-1} \frac{\sum_{i=k+1}^{d} \sum_{j=1}^{k} \bar{a}_{ij} - \sum_{i=1}^{k} \sum_{j=k+1}^{d} \bar{a}_{ij}}{\|u\|_1} V(x + hw_k) \right.$$
$$\left. + \left(\sum_{i=1}^{d} \sum_{j=1, j \neq i}^{d} \alpha_{ij}(x) \ell\left(\frac{\bar{a}_{ij}}{\alpha_{ij}(x)} \right) \right) \frac{h}{\|u\|_1} \right\}, \tag{4.4.4}$$

where $\|u\|_1$ is defined by (4.4.3), and then check if the optimal control u of each candidate infimizer stays in the interior of the positive cone. If it does, then the unique local minimum of the constaint optimization problem is found, and we obtain the global minimum of (4.4.2).

Let $A = \|u\|_1$, we introduce the Lagrange multiplier $\lambda \in \mathbb{R}$ and study

$$\left\{ \sum_{k=1}^{d-1} \frac{\sum_{i=k+1}^{d} \sum_{j=1}^{k} \bar{a}_{ij} - \sum_{i=1}^{k} \sum_{j=k+1}^{d} \bar{a}_{ij}}{A} V(x + hw_k) + \right.$$
$$\left. \left(\sum_{i=1}^{d} \sum_{j=1, j \neq i}^{d} \alpha_{ij}(x) \ell\left(\frac{\bar{a}_{ij}}{\alpha_{ij}(x)} \right) \right) \frac{h}{A} + \frac{\lambda h}{A} \left(\sum_{i=1}^{d} \sum_{j=1, j \neq i}^{d} (i - j) \bar{a}_{ij} - A \right) \right\} \tag{4.4.5}$$

Recall that for $x > 0$, $\ell(x) = x \log x - x + 1$. At each local minimum of (4.4.5), the derivatives with respect to $\{\bar{a}_{ij}\}$ are zero, which leads to:

$$\begin{aligned} -\sum_{k=i}^{j-1} V(x + hw_k) + h \log \frac{\bar{a}_{ij}}{\alpha_{ij}(x)} - (j - i) \lambda h &= 0, \text{ for } 1 \leq i < j \leq d \\ \sum_{k=j}^{i-1} V(x + hw_k) + h \log \frac{\bar{a}_{ij}}{\alpha_{ij}(x)} - (j - i) \lambda h &= 0, \text{ for } 1 \leq j < i \leq d \end{aligned}$$

Thus we can parameterize for some $\lambda \in \mathbb{R}$,

$$\begin{aligned} \bar{a}_{ij} &= \alpha_{ij}(x) e^{(j-i)\lambda + \sum_{k=i}^{j-1} V(x+hw_k)/h}, \text{ for } 1 \leq i < j \leq d \\ \bar{a}_{ij} &= \alpha_{ij}(x) e^{(j-i)\lambda - \sum_{k=j}^{i-1} V(x+hw_k)/h}, \text{ for } 1 \leq j < i \leq d. \end{aligned} \tag{4.4.6}$$

Denote $C=\sum_{i=1}^{d}\sum_{j=1,j\neq i}^{d}\alpha_{ij}\left(x\right)$. Substituting (4.4.6) into (4.4.4), we have

$$\begin{aligned}
& \frac{1}{A}\left[-\sum_{j=1}^{d}\sum_{i=1}^{j-1}\alpha_{ij}\left(x\right)e^{(j-i)\lambda+\sum_{k=i}^{j-1}V(x+hw_k)/h}\left(\sum_{k=i}^{j-1}V\left(x+hw_k\right)\right)\right.\\
& +h\sum_{j=1}^{d}\sum_{i=1}^{j-1}\alpha_{ij}\left(x\right)\ell\left(e^{(j-i)\lambda+\sum_{k=i}^{j-1}V(x+hw_k)/h}\right)\\
& +\sum_{j=1}^{d}\sum_{i=j+1}^{d}\alpha_{ij}\left(x\right)e^{(j-i)\lambda-\sum_{k=j}^{i-1}V(x+hw_k)/h}\left(\sum_{k=j}^{i-1}V\left(x+hw_k\right)\right)\\
& \left.+h\sum_{j=1}^{d}\sum_{i=j+1}^{d}\alpha_{ij}\left(x\right)\ell\left(e^{(j-i)\lambda-\sum_{k=j}^{i-1}V(x+hw_k)/h}\right)\right]\\
= \; & \frac{h}{A}\left[\sum_{j=1}^{d}\sum_{i=1}^{j-1}\left(j-i\right)\lambda\alpha_{ij}\left(x\right)e^{(j-i)\lambda+\sum_{k=i}^{j-1}V(x+hw_k)/h}\right.\\
& -\sum_{j=1}^{d}\sum_{i=1}^{j-1}\alpha_{ij}\left(x\right)e^{(j-i)\lambda+\sum_{k=i}^{j-1}V(x+hw_k)/h}\\
& +\sum_{j=1}^{d}\sum_{i=j+1}^{d}\left(j-i\right)\lambda\alpha_{ij}\left(x\right)e^{(j-i)\lambda-\sum_{k=j}^{i-1}V(x+hw_k)/h}\\
& \left.-\sum_{j=1}^{d}\sum_{i=j+1}^{d}\alpha_{ij}\left(x\right)e^{(j-i)\lambda-\sum_{k=j}^{i-1}V(x+hw_k)/h}+C\right].
\end{aligned}$$

For $b=1,...,d-1$, define

$$\begin{aligned}
K_b &= \sum_{i=1}^{d-b}\sum_{j=1}^{d}\delta_{j-i=b}\alpha_{ij}\left(x\right)e^{\sum_{k=i}^{j-1}V(x+hw_k)/h}\\
K_{-b} &= \sum_{j=1}^{d-b}\sum_{i=1}^{d}\delta_{j-i=-b}\alpha_{ij}\left(x\right)e^{-\sum_{k=j}^{i-1}V(x+hw_k)/h}.
\end{aligned}$$

We can rewrite

$$A=-\sum_{k=1}^{d-1}\left(\sum_{i=1}^{k}\sum_{j=k+1}^{d}\bar{a}_{ij}-\sum_{i=k+1}^{d}\sum_{j=1}^{k}\bar{a}_{ij}\right)=\sum_{b=1}^{d-1}bK_{-b}e^{-b\lambda}-\sum_{b=1}^{d-1}bK_be^{b\lambda}$$

Then (4.4.4) can be further reduced to

$$\begin{aligned} & \frac{h}{A}\left[\sum_{b=1}^{d-1} K_{-b}\left(-b\lambda e^{-b\lambda}-e^{-b\lambda}\right)+\sum_{b=1}^{d-1} K_b\left(b\lambda e^{b\lambda}-e^{b\lambda}\right)+C\right] \\ = \;& h\left(-\lambda-\frac{\sum_{b=1}^{d-1} K_{-b}e^{-b\lambda}+\sum_{b=1}^{d-1} K_b e^{b\lambda}-C}{\sum_{b=1}^{d-1} bK_{-b}e^{-b\lambda}-\sum_{b=1}^{d-1} bK_b e^{b\lambda}}\right). \end{aligned}$$

Minimizing this expression with respect to λ, one looks for λ that satisfies

$$\frac{\left(-\sum_{b=1}^{d-1} b^2K_{-b}e^{-b\lambda}-\sum_{b=1}^{d-1} b^2K_b e^{b\lambda}\right)\left(\sum_{b=1}^{d-1} K_{-b}e^{-b\lambda}+\sum_{b=1}^{d-1} K_b e^{b\lambda}-C\right)}{\left(\sum_{b=1}^{d-1} bK_{-b}e^{-b\lambda}-\sum_{b=1}^{d-1} bK_b e^{b\lambda}\right)^2}=0.$$

Take $\beta=e^{\lambda}$, we then look for the positive solutions β to either

$$\sum_{b=-(d-1)}^{d-1} b^2K_b\beta^b=0$$

or

$$\sum_{b=-(d-1)}^{d-1} K_b\beta^b-C=0. \tag{4.4.7}$$

Recall $\{K_b\}$ are nonnegative, so the first equation has no positive root. We then solve for the $2(d-1)$ roots of (4.4.7). Since the single variable polynomial equations (4.4.7) has only two sign changes between consecutive real coefficients, it follows from Descartes' rule of signs that it has at most two positive real roots. For each positive real root β, one computes $\lambda=\log\beta$, and then obtains the value of $\{\bar{a}_{ij}\}$ by (4.4.6). If the constraints

$$\sum_{i=k+1}^{d}\sum_{j=1}^{k}\bar{a}_{ij}-\sum_{i=1}^{k}\sum_{j=k+1}^{d}\bar{a}_{ij}\geq 0, \quad k=1,...,d-1, \tag{4.4.8}$$

are satisfied, one computes the cost

$$h\left(-\log\beta-\frac{\sum_{b=1}^{d-1}K_{-b}e^{-b\lambda}+\sum_{b=1}^{d-1}K_{b}e^{b\lambda}-C}{\sum_{b=1}^{d-1}bK_{-b}e^{-b\lambda}-\sum_{b=1}^{d-1}bK_{b}e^{b\lambda}}\right)=-h\log\beta,$$

for the local minimizer.

4.4.2 Iterative solver: The narrowest neighborhood structure

Here we present another possible neighborhood structure for the iterations. Recall that in the case of the widest neighborhood structure, for each cone we need to solve an algebraic equation (4.4.7) with degree $2d-2$, which becomes more complicated as dimension grows. For the following narrowest possible neighborhood structure the corresponding equation is always quadratic, but at each $x\in\mathcal{S}^h\backslash B^h$ we need to solve constrained infimization problems for a large number (grows as $O\left(d^2\right)$) of subregions (cones), as d gets large.

Let $\{\tilde{e}_i\}_{i=2}^d$ be defined by $\tilde{e}_i=e_i-e_1$, and we set $\tilde{e}_1=0$. $\{\tilde{e}_2,...,\tilde{e}_d\}$ forms a basis of Δ^{d-1}. For any $x\in\mathcal{S}^h\backslash B^h$ take $\mathcal{V}^*=\mathcal{V}$. Using the same notation as in Section 4.4.1, for $u\in\Delta^{d-1}$ one can write

$$\begin{aligned}-u &= \sum_{i=1}^{d}\sum_{j=1,j\neq i}^{d}\bar{a}_{ij}\left(e_j-e_i\right)=\sum_{i=1}^{d}\sum_{j=1,j\neq i}^{d}\bar{a}_{ij}\left(\tilde{e}_j-\tilde{e}_i\right)\\ &= \sum_{i=2}^{d}\left(\sum_{j=1,j\neq i}^{d}\bar{a}_{ji}-\sum_{j=1,j\neq i}^{d}\bar{a}_{ij}\right)\tilde{e}_i.\end{aligned}$$

Suppose without loss of generality that u belongs to the positive cone spanned by $\tilde{\mathcal{B}}_1=\{\tilde{e}_2,...,\tilde{e}_d\}$. Then $\|u\|_1=-\sum_{i=2}^{d}\left(\sum_{j=1,j\neq i}^{d}\bar{a}_{ji}-\sum_{j=1,j\neq i}^{d}\bar{a}_{ij}\right)=\sum_{i=2}^{d}\bar{a}_{i1}-$

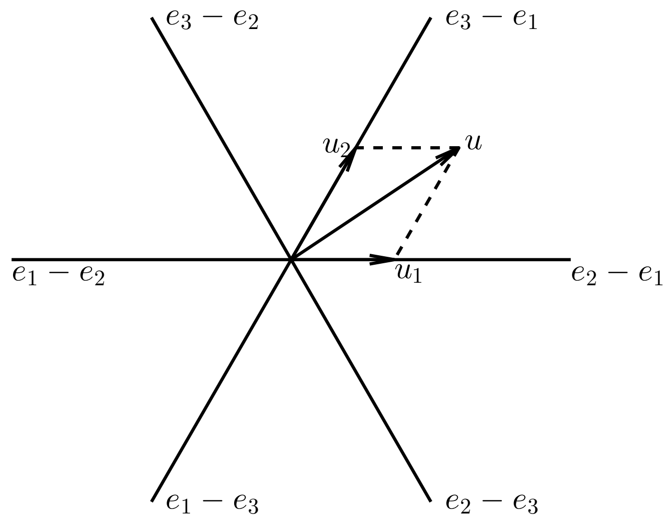

Figure 4.2: Controlled random walks using the narrowest neighborhood structure for $d = 3$

$\sum_{j=2}^{d} \bar{a}_{1j}$.

We construct the controlled random walk such that for $i = 2, ..., d$, take

$$p^h(x, x + h\tilde{e}_i | u) = \frac{\sum_{j=1,j\neq i}^{d} \bar{a}_{ij} - \sum_{j=1,j\neq i}^{d} \bar{a}_{ji}}{\|u\|_1}, \text{ if } x + h\tilde{e}_i \in \mathcal{S}^h,$$

and zero otherwise. Also, set

$$\Delta t^h(u) = \frac{h}{\|u\|_1}.$$

Such a choice of p^h and Δt^h conforms with (4.3.4) and (4.3.5). One then needs to solve the following unconstrained minimization problem

$$\inf_{\{\bar{a}_{ij}\}} \left\{ \sum_{i=2}^{d} \frac{\sum_{j=1,j\neq i}^{d} \bar{a}_{ij} - \sum_{j=1,j\neq i}^{d} \bar{a}_{ji}}{\|u\|_1} V\left(x + h\tilde{e}_i\right) + \left(\sum_{i=1}^{d} \sum_{j=1,j\neq i}^{d} \alpha_{ij}(x)\, \ell \left(\frac{\bar{a}_{ij}}{\alpha_{ij}(x)} \right) \right) \frac{h}{\|u\|_1} \right\}, \tag{4.4.9}$$

and then check if the optimal control of each candidate infimizer stays in the interior of the positive cone. If it does, then the unique local minimum of the constained optimization problem is found, and we obtain the global minimum of (4.4.2).

Letting $B = \|u\|_1$, we introduce the Lagrange multiplier $\lambda \in \mathbb{R}$ and study

$$\left\{ \sum_{i=2}^{d} \frac{\sum_{j=1,j\neq i}^{d} \bar{a}_{ij} - \sum_{j=1,j\neq i}^{d} \bar{a}_{ji}}{B} V\left(x + h\tilde{e}_i\right) + \left(\sum_{i=1}^{d} \sum_{j=1,j\neq i}^{d} \alpha_{ij}(x)\, \ell \left(\frac{\bar{a}_{ij}}{\alpha_{ij}(x)} \right) \right) \frac{h}{B} + \frac{\lambda h}{B} \left[\sum_{i=2}^{d} \bar{a}_{i1} - \sum_{j=2}^{d} \bar{a}_{1j} - B \right] \right\} \tag{4.4.10}$$

At each local minimum of (4.4.9), the derivatives with respect to $\{\bar{a}_{ij}\}$ are zero, which leads to:

$$\begin{aligned}
-V\left(x + h\tilde{e}_j\right) + V\left(x + h\tilde{e}_i\right) + h \log \frac{\bar{a}_{ij}}{\alpha_{ij}(x)} &= 0, \text{ for } 2 \leq i, j \leq d, i \neq j \\
-V\left(x + h\tilde{e}_j\right) + h \log \frac{\bar{a}_{1j}}{\alpha_{1j}(x)} - \lambda h &= 0, \text{ for } j = 2, ..., d \\
V\left(x + h\tilde{e}_i\right) + h \log \frac{\bar{a}_{i1}}{\alpha_{i1}(x)} + \lambda h &= 0, \text{ for } i = 2, ..., d.
\end{aligned}$$

Thus we can parametrize for some $\lambda \in \mathbb{R}$,

$$\begin{aligned}
\bar{a}_{ij} &= \alpha_{ij}(x)\, e^{(V(x+h\tilde{e}_j) - V(x+h\tilde{e}_i))/h}, \text{ for } 2 \leq i, j \leq d, i \neq j \\
\bar{a}_{1j} &= \alpha_{1j}(x)\, e^{\lambda + V(x+h\tilde{e}_j)/h}, \text{ for } j = 2, ..., d \\
\bar{a}_{i1} &= \alpha_{i1}(x)\, e^{-\lambda - V(x+h\tilde{e}_i)/h}, \text{ for } i = 2, ..., d.
\end{aligned} \tag{4.4.11}$$

Substituting (4.4.11) into (4.4.9), we obtain

$$\begin{aligned}
&\frac{1}{B}\left[\sum_{j=2}^{d}\left[-\alpha_{1j}(x)\,e^{\lambda+V(x+h\tilde{e}_j)/h}V(x+h\tilde{e}_j)+h\alpha_{1j}(x)\,\ell\left(e^{\lambda+V(x+h\tilde{e}_j)/h}\right)\right]\right.\\
&+\sum_{i=2}^{d}\left[\alpha_{i1}(x)\,e^{-\lambda-V(x+h\tilde{e}_i)/h}V(x+h\tilde{e}_i)+h\alpha_{i1}(x)\,\ell\left(e^{-\lambda-V(x+h\tilde{e}_i)/h}\right)\right]\\
&+\sum_{i=2}^{d}\sum_{j=2,j\neq i}^{d}\left[\alpha_{ij}(x)\,e^{(V(x+h\tilde{e}_j)-V(x+h\tilde{e}_i))/h}\left(-V(x+h\tilde{e}_j)+V(x+h\tilde{e}_i)\right)\right.\\
&\left.\left.+h\alpha_{ij}(x)\,\ell\left(e^{(V(x+h\tilde{e}_j)-V(x+h\tilde{e}_i))/h}\right)\right]\right].
\end{aligned}$$

Note that

$$\begin{aligned}
& h\alpha_{ij}(x)\,\ell\left(e^{(V(x+h\tilde{e}_j)-V(x+h\tilde{e}_i))/h}\right)\\
= \;& \alpha_{ij}(x)\,e^{(V(x+h\tilde{e}_j)-V(x+h\tilde{e}_i))/h}\left(-V(x+h\tilde{e}_i)+V(x+h\tilde{e}_j)\right)\\
& -h\alpha_{ij}(x)\,e^{(V(x+h\tilde{e}_j)-V(x+h\tilde{e}_i))/h}+h\alpha_{ij}(x).
\end{aligned}$$

Thus if we define

$$\begin{aligned}
K_1 &= \sum_{j=2}^{d}\alpha_{1j}(x)\,e^{V(x+h\tilde{e}_j)/h}, && (4.4.12)\\
K_2 &= \sum_{i=2}^{d}\alpha_{i1}(x)\,e^{-V(x+h\tilde{e}_i)/h} && (4.4.13)\\
C &= \sum_{i=1}^{d}\sum_{j=1,j\neq i}^{d}\alpha_{ij}(x)-\sum_{i=2}^{d}\sum_{j=2,j\neq i}^{d}\alpha_{ij}(x)\,e^{(V(x+h\tilde{e}_j)-V(x+h\tilde{e}_i))/h},
\end{aligned}$$

we can write

$$B=-\sum_{i=2}^{d}\left(\sum_{j=1,j\neq i}^{d}\bar{a}_{ji}-\sum_{j=1,j\neq i}^{d}\bar{a}_{ij}\right)=e^{-\lambda}K_2-e^{\lambda}K_1.$$

And one can further reduce (4.4.9) to

$$\begin{aligned} & \frac{h}{B}\left(-e^{-\lambda}K_2\lambda + e^{\lambda}K_1\lambda - e^{-\lambda}K_2 - e^{\lambda}K_1 + C\right) \\ = \ & h\left(-\lambda - \frac{e^{-\lambda}K_2 + e^{\lambda}K_1 - C}{e^{-\lambda}K_2 - e^{\lambda}K_1}\right). \end{aligned}$$

Minimizing this expression with respect to λ, one looks for λ that solves

$$\frac{\left(e^{-\lambda}K_2 + e^{\lambda}K_1 - C\right)\left(e^{-\lambda}K_2 + e^{\lambda}K_1\right)}{\left(e^{-\lambda}K_2 - e^{\lambda}K_1\right)^2} = 0.$$

Since $e^{-\lambda}K_2 + e^{\lambda}K_1 > 0$, one obtains

$$e^{-\lambda}K_2 + e^{\lambda}K_1 - C = 0,$$

for which one solves

$$\lambda = \log\alpha,$$

with

$$\alpha = \frac{C \pm \sqrt{C^2 - 4K_1K_2}}{2K_1}. \tag{4.4.14}$$

For each root λ, one computes $\{\bar{a}_{ij}\}$ by (4.4.11). If the non-negativity constraints

$$\sum_{j=1,j\neq i}^{d} \bar{a}_{ij} - \sum_{j=1,j\neq i}^{d} \bar{a}_{ji} \geq 0, \text{ for } i = 2, ..., d, \tag{4.4.15}$$

are satisfied, one computes cost

$$h\left(-\log\alpha + \frac{K_2/\alpha + \alpha K_1 - C}{K_2/\alpha - \alpha K_1}\right) = -h\log\alpha, \tag{4.4.16}$$

as the local minimum.

4.4.3 The lower dimensional hyperplanes

To summarize, in the previous two subsections we solve the infimization of (4.4.2) in the relative interior of each cone. The method is to use Lagrange multipliers to first solve the corresponding unconstrained problem, and then test the nonnegativity constraints ((4.4.8) or (4.4.15), depending on the neighborhood structure) for each candidate minimizer. If none of the candidate minimizer satisfies the constraints, we will then need to search lower dimensional boundaries of the cones, until a local minimum is found.

On each lower dimensional boundary, we need to solve a constrained optimization problem, with additional affine constraint added to (4.4.2). If we still use Lagrange multiplier, the calculation of the local minimum is very complicated, especially when the dimension gets further lower (which adds more constraints). In this subsection we discuss solving the constrained optimizations on these lower dimensional hyperplanes, using the results in Section 4.4.1 and 4.4.2.

We show that, for $k = 1, ..., d-2$, to determine candidate minimizers of (4.4.2) on some k dimensional boundaries, it suffices to use jointly the candidate minimizers of (4.4.2) in the relative interior of a cone, and the solution of certain algebraic equations. The advantage of this approach, is that solving the roots of algebraic equations can be implemented much more efficiently than solving constrained optimization problems, especially in high dimensions.

To illustrate, let us assume for now that we use the narrowest neighborhood structure for the interation described in Section 4.4.2 (the other case can be treated in the same way). Suppose we need to solve the constrained minimum on a $d-2$

dimensional boundary of the positive cone generated by $\{\tilde{e}_2, ..., \tilde{e}_d\}$. For $i = 2, ..., d$, denote $u_i = \sum_{j=1, j\neq i}^{d} \bar{a}_{ij} - \sum_{j=1, j\neq i}^{d} \bar{a}_{ji}$, and suppose the boundary is given by $\left\{u \in \Delta^{d-1} : u_2 \geq 0, ..., u_{d-1} \geq 0, u_d = 0\right\}$. We need to solve

$$\begin{aligned} & \inf_{\substack{u\in\Delta^{d-1}:u_2,\ldots,u_{d-1}\geq 0 \\ u_d=0}} \left[\sum_{i=2}^{d} p^h\left(x, x+h\tilde{e}_i|u\right) V^h\left(x+h\tilde{e}_i\right) + L\left(x,-u\right)\Delta t^h\left(u\right)\right] \\ = & \inf_{\substack{u\in\Delta^{d-1}:u_2,\ldots,u_{d-1}\geq 0 \\ u_d=0}} \left[\sum_{i=2}^{d} \frac{u_i}{\left\|u\right\|_1} V^h\left(x+h\tilde{e}_i\right) + L\left(x,-u\right)\frac{h}{\left\|u\right\|_1}\right], \end{aligned} \tag{4.4.17}$$

If we solve the corresponding unconstrained infimization problem, the candidate minimizers are given by (4.4.16).

We now claim the following result. By changing the value of $V^h\left(x+h\tilde{e}_d\right)$, to a particular value V^* (if it exists), then (4.4.17) equals to

$$\inf_{u\in\Delta^{d-1}:u_2,\ldots,u_{d-1}\geq 0} \left[\sum_{i=2}^{d-1} \frac{u_i}{\left\|u\right\|_1} V^h\left(x+h\tilde{e}_i\right) + \frac{u_d}{\left\|u\right\|_1} V^* + L\left(x,-u\right)\frac{h}{\left\|u\right\|_1}\right]. \tag{4.4.18}$$

Note that (4.4.18) can be solved using the results in Section 4.4.2. First obtain the candidate minimizers of the unconstrained problem by (4.4.14), and then check the nonnegative constraints $u_2, ..., u_{d-1} \geq 0$. The value of V^* is determined by jointly solving the algebraic equations

$$\begin{aligned} e^{\lambda} &= \frac{C \pm \sqrt{C^2 - 4K_1K_2}}{2K_1}, \\ \sum_{j=1}^{d-1} \bar{a}_{dj} - \sum_{j=1}^{d-1} \bar{a}_{jd} &= 0, \end{aligned}$$

where $\{\bar{a}_{jd}\left(\lambda, V^*\right)\}$ and $\{\bar{a}_{jd}\left(\lambda, V^*\right)\}$ are specified by (4.4.11), $K_1\left(V^*\right)$ and $K_2\left(V^*\right)$ are given in (4.4.12) and (4.4.13).

To see the claim is true, note that for each such solution V^*, any candidate optimal control u^* to (4.4.18), obtained via the computation in Section 4.4.2, satisfies $u_d^* = 0$ (see (4.4.15)). Therefore the value of (4.4.18) is less than or equal to (4.4.17). The other inequality is straightforward.

This argument works for lower dimensional $(d-3, d-4, ...)$ boundaries as well: one changes a subset of $\left\{V^h\left(x + h\tilde{e}_i\right)\right\}$ to specific values, that are determined by solving a larger systems of algebraic equations jointly, and use the solution to the unconstrained optimization problem in a cone in Section 4.4.1 and 4.4.2. Therefore when search the local minimizer on a lower dimensional boundary, we use the above recipe and convert the problem into solving a system of algebraic equations.

4.5 Numerical Results

In this section, we present numerical results obtained using the Markov chain approximation algorithm described in Section 4.3.

In the actual implementation of the algorithm, iterate the dynamical programming equation by

$$V_{j+1}^h\left(x\right) = \inf_{u \in \Delta^{d-1}} \left[\sum_{v \in \mathcal{V}} p^h\left(x, x + hv|u\right) V_j^h\left(x + hv\right) + L\left(x, -u\right) \Delta t^h\left(u\right)\right]. \quad (4.5.1)$$

Therefore $V_j^h\left(x\right)$ can be interpreted as the minimal cost of the $j-$step optimal control problem with terminal cost V_0^h. If we take the initial data V_0^h, such that for $x \in B^h$, $V_0^h\left(x\right) = V^{(2)}\left(x\right)$, and $V_0^h\left(x\right) = \infty$ for $x \in \mathcal{S}^h \backslash B^h$, then the optimal control is forced to move toward the boundary set B^h quickly, and thus the boundary data can be learned and propagated into the domain quickly. Moreover, the optimal control interpretation implies that $V_j^h\left(x\right)$ is monotone decreasing in j, and converges to the maximal solution of dynamical programming equation (4.3.7).

As is described in Section 4.2, the quadratic approximation $V^{(2)}$ can be obtained by solving the algebraic Riccarti equation (4.2.9) for its Hessian. We use the Matlab build in function to obtain the maximal solution of (4.2.9).

We choose the initial condition to be

$$V_0^h(x) = \begin{cases} V^{(2)}(x) & x \in B^h, \\ M & x \in \mathcal{S}^h \backslash B^h, \end{cases}$$

where $M \in (V, \infty)$ is a very large number, so that the iteration (4.5.1) converges quickly.

For simplicity, we restrict our numerical example below to the case $d = 3$, and embed $\mathcal{S} \subset \mathbb{R}^3$ into $G \doteq \{(x_1, x_2) : x_1 \geq 0, x_2 \geq 0, x_1 + x_2 \leq 1\}$ by taking an affine map. We also focus on the nearest neighborhood structure described in Section 4.4.2.

We use Gauss-Seidel iterative method for our numerical experiments. The Gauss-Seidel iteration goes as follows. During the first iteration, we enforce the value function to be less than or equal to M by letting

$$V_1^h(x) = \inf_{u \in \Delta^{d-1}} \left[\sum_{v \in \mathcal{V}} p^h(x, x + hv|u) V_0^h(x + hv) + L(x, -u) \Delta t^h(u) \right] \wedge M.$$

Suppose the value $\{V_j^h(x)\}$ was assigned to each $x \in \mathcal{S}^h$, at the $(j+1)^{th}$ iteration, we proceed by sweeping through the grid along six possible jump directions in counter-clockwise order (such as sweeping successively along the direction $e_2 - e_1$, $e_3 - e_1$, $e_3 - e_2$, $e_1 - e_2$, $e_1 - e_3$, $e_2 - e_3$). During each sweep, we update sequentially the value at $x \in \mathcal{S}^h \backslash B^h$, by solving (4.4.2), and the new value will be substituted immediately for calculating (4.4.2) at the neighboring points.

Example 4.5.1. *The first example we study is a non-interacting $n-$particle system with 3 states, in the sense that we set $\Gamma_{ij}^{1,n}(\cdot) \equiv 1$ in the generator (2.2.1). The corresponding quasipotential satisfies the Hamilton-Jacobi-Bellman equation*

$$\begin{aligned} \sum_{i=1}^{3} \sum_{j=1, j \neq i}^{3} x_i \exp\left(\frac{\partial V}{\partial x_j} - \frac{\partial V}{\partial x_i}\right) - 2 &= 0, \text{ for } x \in \mathcal{S} \backslash \left\{\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)\right\} \\ V\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) &= 0, \end{aligned}$$

which reduces to

$$\begin{aligned} & x_1 \left(e^{D_{x_1} V} + e^{D_{x_2} V}\right) + x_2 \left(e^{D_{x_2} V - D_{x_1} V} + e^{-D_{x_1} V}\right) \\ & + (1 - x_1 - x_2) \left(e^{D_{x_1} V - D_{x_2} V} + e^{-D_{x_2} V}\right) - 2 = 0 \qquad x \in G \backslash \left(\tfrac{1}{3}, \tfrac{1}{3}\right), \\ & V\left(\tfrac{1}{3}, \tfrac{1}{3}\right) = 0, \end{aligned}$$

by taking the affine map that maps $\mathcal{S}$ to G.

The quasipotential of this problem admits an explicit solution (see also (5.9) of [14])

$$V(x_1, x_2, x_3) = \sum_{i=1}^{3} x_i \log x_i + \log 3.$$

V is a smooth function on any compact subset of int($\mathcal{S}$). Its derivative became unbounded (has logarithmic singularities) when approaching the boundary. As can be seen from the numerical experiment below, the highest error occurs on the boundary.

Let $G_1 = \{x_1 \geq 0.2, x_2 \geq 0.2, x_1 + x_2 \leq 0.8\}$ be a compact subset of G. Let $G^h = G \cap h\mathbb{T}^2$ and $G_1^h = G_1 \cap h\mathbb{T}^2$, where $\mathbb{T}^2$ is the two dimensional triangular lattice. Set $M = 99$. The numerical results are shown in the table below.

Table 4.1: Maximal error table of Example 4.5.1

$n = 1/h$	Max Error in G_1^h	Max Error in G^h
10	0.06	0.1512
20	0.0306	0.1160
30	0.0205	0.0927
40	0.0155	0.0772

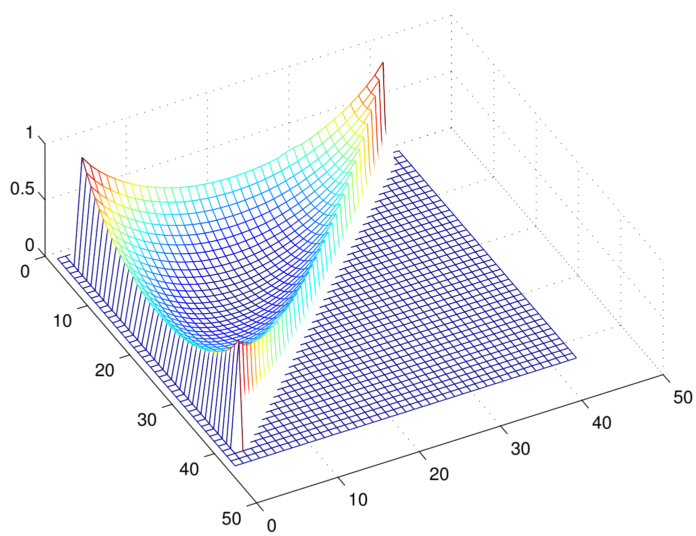

Figure 4.3: The graph of V^h for $h = 1/40$ of Example 4.5.1

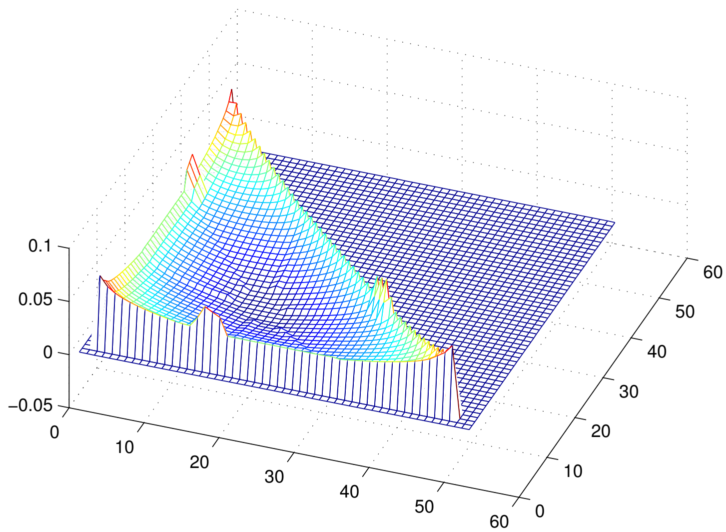

Figure 4.4: Numerical errors for $h = 1/50$ of Example 4.5.1

Example 4.5.2. *We now study the Glauber dynamics of Curie-Weiss-Potts model, described in Example 2.4.2.*

It is known that the Curie-Weiss-Potts model has a phase transition behavior. There exists $\beta_c < \infty$*, called the critical inverse temperature, such that for* $\beta < \beta_c$*, the Markov process defined by (2.2.1) has a unique stationary distribution at* $\delta_{(1/3,1/3,1/3)}$*. For* $d = 3$*,* $\beta_c = 4\log 2$ *(see [8]). In other words,* $(1/3, 1/3, 1/3)$ *is the unique equilibrium point of (4.1.1). In this temperature regime, the associated quasipotential satisfies the Hamilton-Jacobi-Bellman equation*

$$\begin{aligned} \sum_{i=1}^{3}\sum_{j=1,j\neq i}^{3}\frac{x_i\exp\left(-\beta x_i\right)}{\sum_{k=1}^{3}\exp\left(-\beta x_k\right)}\left(\exp\left(\frac{\partial V}{\partial x_j}-\frac{\partial V}{\partial x_i}\right)-1\right) &= 0, x\in\mathcal{S}\backslash\left\{\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)\right\}, \\ V\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right) &= 0. \end{aligned}$$

We study numerical experiments for $\beta = 0.1$*, and take* $x^* = (0.2, 0.2, 0.6)$ *to be a representative point in* $\mathcal{S}$*. Also set* $M = 99$*. Some numerical results are shown in*

Table 4.2: The value of V^h of Example 4.5.2

$n = 1/h$	$V^h(x^*)$
10	0.2157
20	0.1854
30	0.1749
40	0.1697

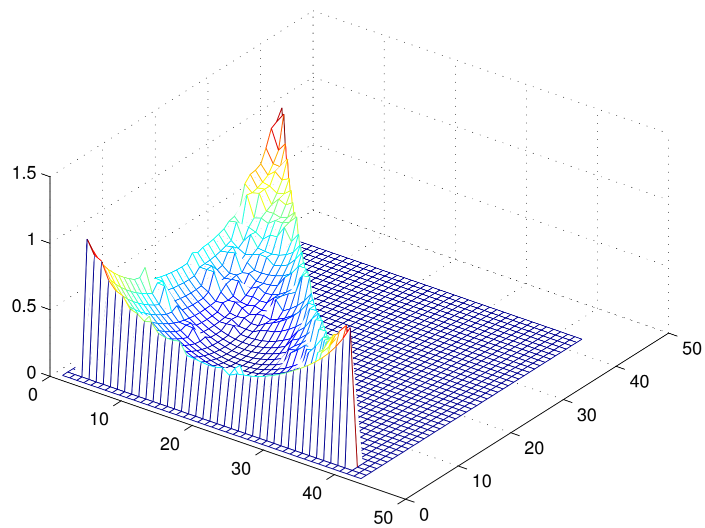

Figure 4.5: The graph of V^h for $h = 1/40$ of Example 4.5.2

the table below.

Appendix A

The Condition of Schwartz and Weiss

In [35] a class of jump Markov process with jump rates vanishing at the boundary of some set $G \subset \mathbb{R}^d$ is studied. We now formulate their conditions in our settings.

Let $G \subset \mathbb{R}^{d-1}$ be the image of $\mathcal{S}$ under the affine map $(x_1, ..., x_{d-1}, 1-\sum_{i\leq d-1} x_i) \mapsto (x_1, ..., x_{d-1})$. Then G is convex and compact. By Lemma 1 of [35] G satisfies an interior cone property with parameters (ε, β): for any $x \in \partial G$, there exists $u \in \mathbb{R}^{d-1}$ such that for each $t \in (0, \varepsilon)$, one has $B(x + tu, \beta t) \subset G$. Also, G can be covered by a finite number of open balls, and the balls can be chosen to be either centered on the boundary, or having no intersection with the boundary. Let $\{B_i\}$ be such a finite cover, Lemma 1 of [35] shows that one can fix the vectors u of the interior cone condition to be some constant u_i in each region B_i. Moreover, there exists γ such that if $d(x, \partial G) < \gamma$, then $d(x + tu_i, \partial G)$ is monotone increasing for $t \in [0, \varepsilon]$.

In [35] they study the jump Markov process with generator

$$\mathcal{L}_n(f)(x) = n \sum_{v \in \mathcal{V}} \lambda_v(x) \left[f\left(x + \frac{1}{n} v\right) - f(x) \right].$$

Recall the definition of $\mathcal{C}\{u_j\}$ in (3.5.3) as the cone generated by $\{u_j\}$, and denote

$$\mathcal{C}_x \doteq \mathcal{C}\{v : v \in \mathcal{V}_x\}.$$

In [35] the following condition is assumed for the proof of LDP.

Condition A.0.3. *The rates and jump directions satisfy the following.*

A. There exists a constant K_λ *such that for all* $v \in \mathcal{V}$, $|\lambda_v(x) - \lambda_v(y)| \leq K_\lambda \|x - y\|$. *Moreover, the rates can be extended to a* $\delta > 0$ *neighborhood of* G*, so that the Lipschitz property still holds.*

B. For each $x \in \partial G$, *there is an* $\varepsilon_1 > 0$ *so that* $y \in \mathcal{C}_x$ *together with* $\|y\| < \varepsilon_1$ *implies* $x + y \in G$.

C. $\lambda_v(x) > 0$ *for all* $v \in \mathcal{V}$ *and* $x \in G^o$. *Moreover,* γ, ε, u_i *of the interior cone condition can be chosen so that*

$$u_i \in \mathcal{C}\left\{v : \inf_{x \in B_i} \lambda_v(x) > \gamma\right\}$$

and if $x \in B_i$, $d(x, \partial G) < \gamma$ *and* $\lambda_v(x) < \gamma$, *then* $\lambda_v(x + tu_i)$ *is strictly monotone increasing for* $t \in (0, \varepsilon)$.

D. $\mathcal{C}\{v : v \in \mathcal{V}\} = \mathbb{R}^{d-1}$.

We provide a simple example of a mean field interacting particle system with $K = 1$, $d = 3$ which does not satisfy the strcitly increasing property in Condition A.0.3.C. The following example is a natural 3 state particle system with a constant rate of one for transitions $1 \to 2$, $2 \to 3$, $3 \to 1$.

Example A.0.4. *Let* $d = 3$ *and* $\{\alpha_{ij}(\cdot)\}$ *specified by* $\alpha_{12}(x) = x_1$, $\alpha_{23}(x) = x_2$, $\alpha_{31}(x) = x_3$, *and zero otherwise. Then* $G = \{(x, y) : x \in [0, 1], y \in [0, 1], x + y \leq 1\}$ *and* $V = \{e_1, -e_2, e_2 - e_1\}$, $\lambda_{e_1}(x) = 1 - x_1 - x_2$, $\lambda_{-e_2}(x) = x_2$, $\lambda_{e_2 - e_1}(x) = x_1$. *For* $x = (1, 0)$, $\{v : \inf_{x \in B_i} \lambda_v(x) > \gamma\} = \{e_2 - e_1\}$, *and by moving along any* $u_i \in \mathcal{C}\{e_2 - e_1\}$, $\lambda_{e_1}(x + tu_i) = 0$, *which contradicts the monotone increasing of* $\lambda_{e_1}(x + tu_i)$ *in* t. *Similar results hold for* $x = (0, 1)$ *and* $x = (0, 0)$.

Appendix B

Proof of Theorem 3.3.6

Recall that h maps a controlled PRM into a controlled process, and is defined in (3.3.6). Recall also the definitions of $\mathcal{A}_b^{\otimes|\mathcal{V}|}$ and $\bar{\mathcal{A}}_b^{\otimes|\mathcal{V}|}$ in Definitions 3.3.5 and 3.3.2, respectively. The claim of Theorem 3.3.6 is essentially that the additional dependence of controls in $\bar{\mathcal{A}}_b^{\otimes|\mathcal{V}|}$ on the "type" of jump is not needed, and that the variational representation is valid with the simpler controls $\mathcal{A}_b^{\otimes|\mathcal{V}|}$. We recall that the controls in $\bar{\mathcal{A}}_b^{\otimes|\mathcal{V}|}$ modulate the intensity of the driving PRM in an s, x and ω dependent fashion, while the controls in $\mathcal{A}_b^{\otimes|\mathcal{V}|}$ multiply the jump rates λ_v^n in an s and ω dependent way.

The proof of Theorem 3.3.6 will follow from Lemma 3.3.4, Lemma B.0.8 and Corollary B.0.7 below. We start with two lemmas which elucidate the relation between elements of $\mathcal{A}_b^{\otimes|\mathcal{V}|}$ and $\bar{\mathcal{A}}_b^{\otimes|\mathcal{V}|}$.

Lemma B.0.5. *There exists a map* $\Theta^n : \bar{\mathcal{A}}_b^{\otimes|\mathcal{V}|} \to \mathcal{A}_b^{\otimes|\mathcal{V}|} \times D\left([0,1] : \mathcal{S}\right) \times \bar{\mathcal{A}}_b^{\otimes|\mathcal{V}|}$ *which takes* $\varphi \in \bar{\mathcal{A}}_b^{\otimes|\mathcal{V}|}$ *into a triple* $(\hat{\alpha}^n, \hat{\mu}^n, \hat{\varphi}^n)$*, such that for any* $v \in \mathcal{V}$

1. $\hat{\alpha}_v^n\left(s\right) = \int_0^{\lambda_v^n(\hat{\mu}^n(s))} \varphi_v\left(s,x\right) dx$

2. $\hat{\varphi}_v^n\left(s,x\right) = \frac{\hat{\alpha}_v^n(s)}{\lambda_v^n(\hat{\mu}^n(s))} 1_{[0,\lambda_v^n(\hat{\mu}^n(s))]}(x) + 1_{[0,\lambda_v^n(\hat{\mu}^n(s))]^c}(x)$

3. $\hat{\mu}^n = h\left(\frac{1}{n} N^{n\hat{\varphi}^n}, \mu^n\left(0\right), \lambda^n\right)$.

Note that given any control $\varphi \in \bar{\mathcal{A}}_b^{\otimes|\mathcal{V}|}$, this lemma identifies a structurally simpler control $\hat{\varphi}^n \in \bar{\mathcal{A}}_b^{\otimes|\mathcal{V}|}$.

Proof. We prove the claim by a recursive construction.

1. To begin the recursion set $t_0 = 0$, and given any $\varphi \in \bar{\mathcal{A}}_b^{\otimes|\mathcal{V}|}$, define for $s \geq t_0$

and $v \in \mathcal{V}$,

$$\begin{aligned}
\hat{\mu}^{n,0}(s) &= \mu^n(0), \\
\hat{\alpha}_v^{n,0}(s) &= \int_0^{\lambda_v^n(\hat{\mu}^{n,0}(s))} \varphi_v(s,x)\,dx, \\
\hat{\varphi}_v^{n,0}(s,x) &= \frac{\hat{\alpha}_v^{n,0}(s)}{\lambda_v^n(\hat{\mu}^{n,0}(s))} 1_{[0,\lambda_v^n(\hat{\mu}^{n,0}(s))]}(x) + 1_{[0,\lambda_v^n(\hat{\mu}^{n,0}(s))]^c}(x).
\end{aligned}$$

In other words, for $s > 0$ and x inside the compact set $[0, \lambda_v^n(\hat{\mu}^{n,0}(s))]$, we set $\hat{\varphi}_v^{n,0}$ to be the average of $\varphi_v(s,\cdot)$ over the set, while for x in the complement we set $\hat{\varphi}_v^{n,0} = 1$. We see that by construction $\|\hat{\varphi}_v^{n,0}\|_\infty \leq \|\varphi_v\|_\infty \vee 1$.

2. Assume now that for some $k \in \mathbb{N}_0$, t_k is well defined, $(\{\hat{\varphi}_v^{n,k}(s)\}, \{\hat{\alpha}_v^{n,k}(s)\}, \{\hat{\mu}^{n,k}(s)\})$ is well defined for $s \in [0,1]$, and

$$\left\|\hat{\varphi}_v^{n,k}\right\|_{\infty,[t_k,\infty)} \doteq \sup_{(s,x)\in[t_k,\infty)\times\mathbb{R}} \left|\hat{\varphi}_v^{n,k}(s,x)\right| \leq \|\varphi_v\|_\infty \vee 1.$$

For any $t \geq t_k$ and $v \in \mathcal{V}$, define

$$\hat{B}_{k,v}(t) = \left\{(s,x,r) : s \in [t_k, t], x \in \left[0, \lambda_v^n(\hat{\mu}^{n,k}(s))\right], r \in \left[0, \hat{\varphi}_v^{n,k}(s,x)\right]\right\}$$

and

$$t_{k+1} = \inf\left\{t > t_k \text{ such that for some } v \in \mathcal{V}, \bar{N}_v^n(\hat{B}_{k,v}(t)) > 0\right\} \wedge 1.$$

We define $\hat{\mu}^{n,k+1}$ on $[0,1]$ by setting $\hat{\mu}^{n,k+1}(s) = \hat{\mu}^{n,k}(s)$ for $s \in [0, t_{k+1}]$. At t_{k+1}, update $\hat{\mu}^{n,k+1}$ by setting

$$\begin{aligned}
\hat{\mu}^{n,k+1}(t_{k+1}) &= \hat{\mu}^{n,k}(t_k) \\
&+ \sum_{v\in\mathcal{V}} v \int_{[t_k,t_{k+1})} \int_{\mathcal{Y}} 1_{\left[0,\lambda_v^n\left(\hat{\mu}^{n,k}(s)\right)\right]}(x) \int_{[0,\infty)} 1_{[0,\hat{\varphi}_v^{n,k}(s,x)]}(r) \frac{1}{n} \bar{N}_v^n(dsdxdr),
\end{aligned}$$

and set $\hat{\mu}^{n,k+1}(s) = \hat{\mu}^{n,k+1}(t_{k+1})$ for $s \in [t_{k+1}, 1]$. Define

$$\begin{aligned}\hat{\alpha}_v^{n,k+1}(s) &= \int_0^{\lambda_v^n(\hat{\mu}^{n,k+1}(s))} \varphi_v(s,x)\,dx,\\ \hat{\varphi}_v^{n,k+1}(s,x) &= \frac{\hat{\alpha}_v^{n,k+1}(s)}{\lambda_v^n(\hat{\mu}^{n,k+1}(s))} 1_{[0,\lambda_v^n(\hat{\mu}^{n,k+1}(s))]}(x) + 1_{[0,\lambda_v^n(\hat{\mu}^{n,k+1}(s))]^c}(x).\end{aligned}$$

We also have $\left\|\hat{\varphi}_v^{n,k+1}\right\|_{\infty,[t_{k+1},\infty)} \leq \|\varphi_v\|_\infty \vee 1$.

3. Recall M' defined as in (3.3.5). Since $\bar{N}_v^n$ has a.s. finitely many atoms on $[0,1] \times [0, M'] \times [0, \|\varphi_v\|_\infty]$, the construction will produce functions defined on $[0,1]$ in $L < \infty$ steps. Then set

$$\hat{\mu}^n(s) = \hat{\mu}^{n,L}(s),\ \hat{\alpha}_v^n(s) = \hat{\alpha}_v^{n,L}(s),\ \hat{\varphi}_v^n(s) = \hat{\varphi}_v^{n,L}(s),\ \text{if } s \in [0,1].$$

Then $\hat{\varphi}^n \in \bar{\mathcal{A}}_b^{\otimes|\mathcal{V}|}$. Furthermore, by construction

$$\hat{\mu}^n = h\left(\frac{1}{n} N^{n\hat{\varphi}^n}, \mu^n(0), \lambda^n\right).$$

□

The next lemma shows that from controls in $\mathcal{A}_b^{\otimes|\mathcal{V}|}$ we can produce corresponding controls in $\bar{\mathcal{A}}^{\otimes|\mathcal{V}|}$.

Lemma B.0.6. *There exists a map $\Xi^n : \mathcal{A}_b^{\otimes|\mathcal{V}|} \to D([0,1] : \mathcal{S}) \times \bar{\mathcal{A}}^{\otimes|\mathcal{V}|}$ which takes $\bar{\alpha} \in \mathcal{A}_b^{\otimes|\mathcal{V}|}$ into a pair $(\bar{\mu}^n, \bar{\varphi})$, such that $\bar{\mu}^n = h\left(\frac{1}{n}N^{n\bar{\varphi}}, \mu^n(0), \lambda^n\right)$, where for $v \in \mathcal{V}$, $\bar{\varphi}_v(s,x) = \frac{\bar{\alpha}_v(s)}{\lambda_v^n(\bar{\mu}^n(s))} 1_{[0,\lambda_v^n(\bar{\mu}^n(s))]}(x) + 1_{[0,\lambda_v^n(\bar{\mu}^n(s))]^c}(x)$.*

Proof. 1. Define $t_0 = 0$ and for any $\bar{\alpha} \in \mathcal{A}_b^{\otimes|\mathcal{V}|}$, $s \geq t_0$ and $v \in \mathcal{V}$, define

$$\begin{aligned}
\bar{\mu}^{n,0}(s) &= \mu^n(0), \\
\bar{\varphi}_v^0(s,x) &= \frac{\bar{\alpha}_v(s)}{\lambda_v^n(\bar{\mu}^{n,0}(s))} 1_{[0,\lambda_v^n(\bar{\mu}^{n,0}(s))]}(x) + 1_{[0,\lambda_v^n(\bar{\mu}^{n,0}(s))]^c}(x).
\end{aligned}$$

2. Assume now that for some $k \in \mathbb{N}$, t_k is well defined, and that $\left(\bar{\mu}^{n,k}(s), \left\{\bar{\varphi}_v^k(s)\right\}\right)$ is well defined for $s \geq t_k$. For any $t \geq t_k$, define

$$\bar{A}_{k,v}(t) = \left\{(s,x,r) : s \in [t_k, t], x \in \left[0, \lambda_v^n\left(\bar{\mu}^{n,k}(s)\right)\right], r \in \left[0, \bar{\varphi}_v^k(s,x)\right]\right\}$$

and

$$t_{k+1} = \inf\left\{t > t_k \text{ such that for some } v \in \mathcal{V}, \bar{N}_v^n\left(\bar{A}_{k,v}(t)\right) > 0\right\} \wedge 1.$$

We define $\bar{\mu}^{n,k+1}$ on $[0,1]$ by setting $\bar{\mu}^{n,k+1}(s) = \bar{\mu}^{n,k}(s)$ for $s \in [0, t_{k+1}]$. At t_{k+1}, update $\bar{\mu}^{n,k+1}$ by

$$\begin{aligned}
\bar{\mu}^{n,k+1}(t_{k+1}) &= \bar{\mu}^{n,k}(t_k) \\
&+ \sum_{v \in \mathcal{V}} v \int_{[t_k,t_{k+1})} \int_{\mathcal{Y}} 1_{\left[0,\lambda_v^n\left(\bar{\mu}^{n,k}(s)\right)\right]}(x) \int_{[0,\infty)} 1_{[0,\bar{\varphi}_v^k(s,x)]}(r) \frac{1}{n} \bar{N}_v^n(dsdxdr),
\end{aligned}$$

and set $\bar{\mu}^{n,k+1}(s) = \bar{\mu}^{n,k+1}(t_{k+1})$ for $s \in [t_{k+1}, 1]$. Define

$$\bar{\varphi}_v^{k+1}(s,x) = \frac{\bar{\alpha}_v(s)}{\lambda_v^n\left(\bar{\mu}^{n,k+1}(s)\right)} 1_{\left[0,\lambda_v^n\left(\bar{\mu}^{n,k+1}(s)\right)\right]}(x) + 1_{\left[0,\lambda_v^n\left(\bar{\mu}^{n,k+1}(s)\right)\right]^c}(x).$$

3. Since $\bar{N}_v^n$ has a.s. finitely many atoms on $[0,1] \times [0, \|\bar{\alpha}_v\|_\infty]$, the construction

will produce functions defined on $[0,1]$ in $L<\infty$ steps. Then set

$$\bar{\mu}^n(s)=\bar{\mu}^{n,L}(s)\,,\ \bar{\varphi}_v(s)=\bar{\varphi}_v^L(s)\,,\ s\in[0,1].$$

Note that

$$\bar{\varphi}_v(s,x)=\frac{\bar{\alpha}_v(s)}{\lambda_v^n(\bar{\mu}^n(s))}1_{[0,\lambda_v^n(\bar{\mu}^n(s))]}(x)+1_{[0,\lambda_v^n(\bar{\mu}^n(s))]^c}(x)$$

and $\bar{\mu}^n$ satisfies

$$\bar{\mu}^n=h\left(\frac{1}{n}N^{n\bar{\varphi}},\mu^n(0)\,,\lambda^n\right).$$

□

The next result is a corollary to the construction in Lemma B.0.6.

Corollary B.0.7. *For any $\{\bar{\alpha}_v\}\in\mathcal{A}_b^{\otimes|\mathcal{V}|}$ given and $t\in[0,1]$, $\Xi_1^n(\bar{\alpha})(t)$ (defined in Lemma B.0.6) has the same distribution as $\Lambda^n(\bar{\alpha},\mu^n(0))(t)$, where Λ^n is as defined in (3.3.9).*

Proof. Recall that $\bar{\mu}^n=\Xi_1^n(\bar{\alpha})$. We have $\Xi_1^n(\bar{\alpha})(0)=\mu^n(0)$. Given $s\in[0,1]$, the total jump intensity of $\bar{\mu}^n(s)$ in the direction v is

$$\int_0^{\lambda_v^n(\bar{\mu}^n(s))}\bar{\varphi}_v(s,x)\,dx=\int_0^{\lambda_v^n(\bar{\mu}^n(s))}\frac{\bar{\alpha}_v(s)}{\lambda_v^n(\bar{\mu}^n(s))}dx=\bar{\alpha}_v(s)$$

which is the same as that of $\Lambda^n(\bar{\alpha},\mu^n(0))(s)$. □

Lemma B.0.8. *Let $\mathcal{A}_b$, $\bar{\mathcal{A}}_b$ and $\bar{\mathcal{A}}$ be as defined in Definition 3.3.5, Definition 3.3.2* *and Definition 3.3.1 respectively. Then for $F \in M_b(\mathcal{S})$,*

$$\begin{aligned}
&\inf_{\varphi \in \bar{\mathcal{A}}_b^{\otimes|\mathcal{V}|}} \bar{\mathbb{E}}\left[\sum_{v\in\mathcal{V}} L_1(\varphi_v) + F(\bar{\mu}^n) : \bar{\mu}^n = h\left(\frac{1}{n}N^{n\varphi}, \mu^n(0), \lambda^n\right)\right] \\
&\quad = \inf_{\varphi \in \bar{\mathcal{A}}^{\otimes|\mathcal{V}|}} \bar{\mathbb{E}}\left[\sum_{v\in\mathcal{V}} L_1(\varphi_v) + F(\bar{\mu}^n) : \bar{\mu}^n = h\left(\frac{1}{n}N^{n\varphi}, \mu^n(0), \lambda^n\right)\right] \\
&\quad = \inf_{\bar{\alpha} \in \mathcal{A}_b^{\otimes|\mathcal{V}|}} \bar{\mathbb{E}}\left[\sum_{v\in\mathcal{V}} \int_0^1 \lambda_v^n(\bar{\mu}^n(t))\, \ell\left(\frac{\bar{\alpha}_v(t)}{\lambda_v^n(\bar{\mu}^n(t))}\right) dt + F(\bar{\mu}^n) : \bar{\mu}^n = \Xi_1^n(\bar{\alpha})\right],
\end{aligned}$$

where Ξ^n is as defined in Lemma B.0.6.

Proof. The first equality is a consequence of Theorem 2.4 of [5]. To prove the rest of the claim, for $\bar{\alpha} \in \mathcal{A}_b^{\otimes|\mathcal{V}|}$ fixed, let $(\bar{\mu}^n, \bar{\varphi}) = \Xi^n(\bar{\alpha})$. Then by definition $\bar{\varphi} \in \bar{\mathcal{A}}^{\otimes|\mathcal{V}|}$, and

$$\sum_{v\in\mathcal{V}} \int_0^\infty \int_0^1 \ell(\bar{\varphi}_v(t,x))\, dt dx = \sum_{v\in\mathcal{V}} \int_0^1 \lambda_v^n(\bar{\mu}^n(t))\, \ell\left(\frac{\bar{\alpha}_v(t)}{\lambda_v^n(\bar{\mu}^n(t))}\right) dt.$$

Now it follows from Lemma B.0.6 that

$$\begin{aligned}
&\inf_{\varphi \in \bar{\mathcal{A}}^{\otimes|\mathcal{V}|}} \bar{\mathbb{E}}\left[\sum_{v\in\mathcal{V}} L_1(\varphi_v) + F(\bar{\mu}^n) : \bar{\mu}^n = h\left(\frac{1}{n}N^{n\varphi}, \mu^n(0), \lambda^n\right)\right] \\
&\quad \le \bar{\mathbb{E}}\left[\sum_{v\in\mathcal{V}} L_1(\bar{\varphi}_v) + F \circ h\left(\frac{1}{n}N^{n\bar{\varphi}}, \mu^n(0), \lambda^n\right)\right] \\
&\quad = \bar{\mathbb{E}}\left[\sum_{v\in\mathcal{V}} \int_0^1 \lambda_v^n(\bar{\mu}^n(t))\, \ell\left(\frac{\bar{\alpha}_v(t)}{\lambda_v^n(\bar{\mu}^n(t))}\right) dt + F(\bar{\mu}^n)\right].
\end{aligned}$$

The reverse inequality is proved by a convexity argument. Recall the definition of Θ^n given in Lemma B.0.5. For given $\varphi \in \bar{\mathcal{A}}_b^{\otimes|\mathcal{V}|}$, let $(\bar{\alpha}, \bar{\mu}^n) \doteq (\Theta_1^n(\varphi), \Theta_2^n(\varphi))$. Then

$\bar{\alpha} \in \mathcal{A}_b^{\otimes|\mathcal{V}|}$. By convexity of $\ell(\cdot)$ and Jensen's inequality,

$$\begin{aligned}
&\int_0^\infty \int_0^1 \ell\left(\varphi_v\left(s,x\right)\right) dsdx \\
&\quad \geq \int_0^1 \lambda_v^n\left(\bar{\mu}^n(s)\right)\left(\frac{1}{\lambda_v^n\left(\bar{\mu}^n(s)\right)}\int_0^{\lambda_v^n(\bar{\mu}^n(s))} \ell\left(\varphi_v\left(s,x\right)\right) dx\right) ds \\
&\quad \geq \int_0^1 \lambda_v^n\left(\bar{\mu}^n(s)\right) \ell\left(\frac{1}{\lambda_v^n\left(\bar{\mu}^n(s)\right)}\int_0^{\lambda_v^n(\bar{\mu}^n(s))} \varphi_v\left(s,x\right) dx\right) ds \\
&\quad = \int_0^1 \lambda_v^n\left(\bar{\mu}^n(s)\right) \ell\left(\frac{\bar{\alpha}_v\left(s\right)}{\lambda_v^n\left(\bar{\mu}^n(s)\right)}\right) ds
\end{aligned}$$

Summing over $v \in \mathcal{V}$, applying Lemma B.0.6 and infimizing over $\bar{\alpha} \in \mathcal{A}_b^{\otimes|\mathcal{V}|}$ we obtain the desired result. $\square$

Appendix C

Proof of Theorem 3.4.1

Proof. The idea is to use uniformization to represent $X^n - Y^n$ as a jump process with (well controlled) small jump rates. Since $\lambda_v^n(\cdot), \lambda_v(\cdot)$ are bounded by some constant C, let $\gamma \doteq |\mathcal{V}|(C+1)$. Also denote $q(x) = \sum_{v\in\mathcal{V}} \lambda_v(x)$ and $q^n(x) = \sum_{v\in\mathcal{V}} \lambda_v^n(x)$. We then define the uniformized transition matrices on $\mathcal{S}_n$

$$\begin{aligned} P\left(x, x+\frac{1}{n}v\right) &= \frac{\lambda_v(x)}{\gamma}, P(x,x) = 1 - \frac{q(x)}{\gamma} \\ P^n\left(x, x+\frac{1}{n}v\right) &= \frac{\lambda_v^n(x)}{\gamma}, P^n(x,x) = 1 - \frac{q^n(x)}{\gamma}. \end{aligned}$$

And let $\tilde{X}^n, \tilde{Y}^n$ be the Markov chain on $\mathcal{S}_n$ starting at x, and with transition matrix P, P^n respectively. Let N_t be a Poisson process with rate γ, and is independent of $\tilde{X}_t^n$ and $\tilde{Y}_t^n$. Then $\tilde{X}_{N_t}^n, \tilde{Y}_{N_t}^n$ has the same distribution as X_t^n, Y_t^n, respectively. It then suffices to couple $\tilde{X}^n$ and $\tilde{Y}^n$ so that their displacement is a Markov chain with small transition rate.

We can construct a coupling $\left(\bar{X}^n, \bar{Y}^n\right)$ of $\tilde{X}^n$ and $\tilde{Y}^n$ as follows. We define a Markov chain on $\mathcal{S}_n \times \mathcal{S}_n$ that starts at (x,x) and with transition probability that for each $v \in \mathcal{V}$,

$$\begin{aligned} (x,y) &\to \left(x+\frac{1}{n}v, y+\frac{1}{n}v\right) && \text{with probability } \min\{\lambda_v(x), \lambda_v^n(y)\}/\gamma \\ (x,y) &\to \left(x+\frac{1}{n}v, y\right) && \text{with probability } (\lambda_v(x) - \lambda_v^n(y))^+/\gamma \\ (x,y) &\to \left(x, y+\frac{1}{n}v\right) && \text{with probability } (\lambda_v(x) - \lambda_v^n(y))^-/\gamma \end{aligned}$$

and stay at (x,y) otherwise. Let $\bar{Z}^n = \bar{X}^n - \bar{Y}^n$, then $\bar{Z}^n(0) = 0$, and it jumps in some direction $v \in \mathcal{V}$ with probability at most $|\lambda_v(x) - \lambda_v^n(y)|/\gamma \le (C\|x-y\| + O(1/n))/\gamma$, if the current position is $x - y$. Thus $X_t^n - Y_t^n$ has the same distribution as $Z_t^n \doteq \bar{Z}_{N_t}^n$, which is a Markov process starts at 0 and jumps in some direction $v \in \mathcal{V}$ with rates at most $C\|x-y\| +$

$O(1/n)$ for some $C < \infty$. And it sufficess to get an superexponential bound for $\mathbb{P}\left(\sup_{t\in[0,1]} \|Z^n(t)\| > \delta\right)$.

If Z^n were on one dimensional lattice $\frac{1}{n}\mathbb{N}$ this bound is implied by Theorem 3.6.1. Let W^n be a birth process on $\frac{1}{n}\mathbb{N}$ which jumps to the right neighbor with rate $q_{x,x+1/n} = Cx + C_2/n$, for some C_2 large enough, and stay at x otherwise. By the comparison principle of ODE, it follows that $\mathbb{P}_{1d}(Z^n(t) \geq c) \leq \mathbb{P}_{1d}(W^n(t) \geq c)$ for any $c > 0$ and $t \in [0,1]$. Thus

$$\begin{aligned}
\mathbb{P}_{1d}\left(\sup_{t\in[0,1]} |Z^n(t)| > \delta\right) &\leq \mathbb{P}_{1d}\left(\sup_{t\in[0,1]} |W^n(t)| > \delta\right) \leq \mathbb{P}_{1d}(|W^n(1)| > \delta) \\
&\leq \frac{1}{(\delta n)!}\prod_{j=1}^{\delta n}(Cj/n + C_2/n) \\
&= \left(\frac{C}{n}\right)^{\delta n}\binom{C_2/C + \delta n}{C_2/C} \\
&\leq \left(\frac{C}{n}\right)^{\delta n}(C_2/C + \delta n)^{C_2/C}
\end{aligned}$$

which goes to 0 superexponentially as $n \to \infty$.

The general result follows from one dimensional path counting. For n sufficiently large, we have

$$
\begin{aligned}
& \mathbb{P}\left(\sup_{t\in[0,1]} \|Z^n(t)\| > \delta\right) \\
\leq\ & \mathbb{P}\left(Z^n \text{ jumps at least } \delta n \text{ lattice steps in } [0,1]\right) \\
\leq\ & \sum_{k=\delta n}^{\infty} \mathbb{P}\left(Z^n \text{ jumps } k \text{ lattice steps in } [0,1]\right) \\
\leq\ & \sum_{k=\delta n}^{\infty} \#\left\{\text{lattice path in } \mathcal{S}_n \text{ with length } k\right\} \cdot \mathbb{P}_{1d}\left(\sup_{t\in[0,1]} |Z^n(t)| > k/n\right) \\
\leq\ & \sum_{k=\delta n}^{\infty} (2d)^k \left[C/n\right]^k \left(k + C_2/C\right)^{C_2/C} \\
\leq\ & \sum_{k=\delta n}^{\infty} (2d)^k \left[C/n\right]^{k/2} \\
=\ & \left[4d^2C/n\right]^{\delta n/2} \frac{1}{1-4d^2C/n},
\end{aligned}
$$

where the last inequality follows from the fact that $\left[C/n\right]^{k/2}\left(k + C_2/C\right)^{C_2/C} < 1$ for n sufficiently large. Therefore

$$
\begin{aligned}
& -\frac{1}{n}\log\mathbb{P}\left(\sup_{t\in[0,1]} \|Z^n(t)\| > \delta\right) \\
\geq\ & -\frac{1}{n}\left(\frac{\delta n}{2}\log\left[4d^2C/n\right] + C_3\right) \\
\geq\ & \frac{\delta}{2}\log n + C_4,
\end{aligned}
$$

for some $C_3, C_4 > 0$. The expression goes to ∞ as $n \to \infty$. $\square$

Bibliography

[1] Nelson Antunes, Christine Fricker, Philippe Robert, and Danielle Tibi. Analysis of loss networks with routing. *The Annals of Applied Probability*, 16(4):2007–2026, 2006.

[2] Abraham Berman and Robert J Plemmons. *Nonnegative matrices*. SIAM, 1979.

[3] Vivek S Borkar and Rajesh Sundaresan. Asymptotics of the invariant measure in mean field models with jumps. In *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on*, pages 1258–1263. IEEE, 2011.

[4] Michelle Boué and Paul Dupuis. Markov chain approximations for deterministic control problems with affine dynamics and quadratic cost in the control. *SIAM Journal on Numerical Analysis*, 36(3):667–695, 1999.

[5] Amarjit Budhiraja, Jiang Chen, and Paul Dupuis. Large deviations for stochastic partial differential equations driven by a Poisson random measure. *Stochastic Processes and their Applications*, 123:523–560.

[6] Amarjit Budhiraja, Paul Dupuis, and Vasileios Maroulas. Variational representations for continuous time processes. 47(3):725–747, 2011.

[7] Amarjit Budhiraja, Paul Dupuis, and Kavita Ramanan. *in preparation*.

[8] Marius Costeniuc, Richard S Ellis, and Hugo Touchette. Complete analysis of phase transitions and ensemble equivalence for the curie–weiss–potts model. *Journal of mathematical physics*, 46(6):063301, 2005.

[9] Donald A Dawson and Jürgen Gärtner. Large deviations from the Mckean-Vlasov limit for weakly interacting diffusions. *Stochastics: An International Journal of Probability and Stochastic Processes*, 20(4):247–308, 1987.

[10] Martin V Day and Thomas A Darden. Some regularity results on the Ventcel-Freidlin quasi-potential function. *Applied Mathematics and Optimization*, 13(1):259–282, 1985.

[11] Amir Dembo and Andrea Montanari. Gibbs measures and phase transitions on sparse random graphs. *Brazilian Journal of Probability and Statistics*, 24(2):137–211, 2010.

[12] Paul Dupuis and Richard S Ellis. *A weak convergence approach to the theory of large deviations*. John Wiley & Sons, 1997.

[13] Paul Dupuis, Richard S Ellis, and Alan Weiss. Large deviations for Markov processes with discontinuous statistics, I: General upper bounds. *The Annals of Probability*, 19(3):1280–1297, 1991.

[14] Paul Dupuis and Markus Fischer. On the construction of Lyapunov functions for nonlinear Markov processes via relative entropy. *submitted for publication*, 2011.

[15] Paul Dupuis, Carl Nuzman, and Phil Whiting. Large deviation asymptotics for occupancy problems. *The Annals of Probability*, 32(3B):2765–2818, 2004.

[16] Paul Dupuis and John Oliensis. An optimal control formulation and related numerical methods for a problem in shape reconstruction. *The Annals of Applied Probability*, 4:287–346, 1994.

[17] Paul Dupuis, Kavita Ramanan, and Wei Wu. Markov chain approximations for quasipotentials of finite state mean-field interacting particle systems. *in preparation*.

[18] Rick Durrett. Ten lectures on particle systems. In *Lectures on Probability Theory*, pages 97–201. Springer, 1995.

[19] Wendell H Fleming and H Mete Soner. *Controlled Markov processes and viscosity solutions*. Springer-Verlag (New York), 1993.

[20] Mark Freidlin and Alexander D Wentzell. *Random perturbations of dynamical systems*. Springer, 2012.

[21] RJ Gibbens, PJ Hunt, and FP Kelly. Bistability in communication networks. *Disorder in physical systems*, pages 113–128, 1990.

[22] Diogo A Gomes, Joana Mohr, and Rafael Rigao Souza. Discrete time, finite state space mean field games. *Journal de mathématiques pures et appliquées*, 93(3):308–328, 2010.

[23] C Graham and S Méléard. A large deviation principle for a large star-shaped loss network with links of capacity one. *Markov Processes and Related Fields*, 3(4):475–492, 1997.

[24] Carl Graham and Neil O'Connell. Large deviations at equilibrium for a large star-shaped loss network. *The Annals of Applied Probability*, 10(1):104–122, 2000.

[25] Mark Kac. Foundations of kinetic theory. In *Berkeley symp. on Math. Stat. and Prob.*, volume 3, pages 171–197. Univ. of Calif. Press, 1956.

[26] Vassili N Kolokoltsov. *Nonlinear Markov processes and kinetic equations*. Cambridge University Press London, 2010.

[27] Thomas G Kurtz. Solutions of ordinary differential equations as limits of pure jump Markov processes. *Journal of Applied Probability*, 7(1):49–58, 1970.

[28] Harold Kushner. *Probability methods for approximations in stochastic control and for elliptic equations*, volume 129. Academic Press New York, 1977.

[29] Harold J Kushner and Paul Dupuis. *Numerical methods for stochastic control problems in continuous time*. Springer, 2001.

[30] Christian Leonard. Large deviations for long range interacting particle systems with jumps. *Annales de l'IHP Probabilités et statistiques*, 31(2):289–323, 1995.

[31] David A Levin, Malwina J Luczak, and Yuval Peres. Glauber dynamics for the mean-field Ising model: cut-off, critical power law, and metastability. *Probability theory and related fields*, 146(1-2):223–265, 2010.

[32] Horacio J Marquez. *Nonlinear control systems: analysis and design*. John Wiley, 2003.

[33] Fabio Martinelli. Lectures on Glauber dynamics for discrete spin models. In *Lectures on probability theory and statistics*, pages 93–191. Springer, 1999.

[34] Karl Oelschlager. A martingale approach to the law of large numbers for weakly interacting stochastic processes. *The Annals of Probability*, 12(2):458–479, 1984.

[35] Adam Shwartz and Alan Weiss. Large deviations with diminishing rates. *Mathematics of Operations Research*, 30(2):281–310, 2005.

[36] Adam Shwartz, Alan Weiss, and Robert J Vanderbei. *Large deviations for performance analysis*. Chapman and Hall, 1995.

[37] Eduardo D Sontag. *Mathematical control theory: deterministic finite dimensional systems*. Springer, 1998.

[38] Danielle Tibi. Metastability in communication networks. *arXiv preprint arXiv:1002.0796*, 2010.

[39] ACD van Enter, Roberto Fernández, Frank den Hollander, and Frank Redig. A large-deviation view on dynamical Gibbs-non-Gibbs transitions. *Moscow Math. J*, 10:687–711, 2010.

[40] SR Srinivasa Varadhan. *Large deviations and applications*. SIAM, 1984.